PATTERN AND PROCESS IN A MOSAIC HYBRID ZONE: FROM PHENOTYPE TO GENOTYPE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Erica Lee Larson

January 2013

PATTERN AND PROCESS IN A MOSAIC HYBRID ZONE:

FROM PHENOTYPE TO GENOTYPE

Erica Lee Larson, Ph. D.

Cornell University 2013

Speciation is the process of a single lineage splitting into two or more daughter lineages over time. The evolution of barriers to gene exchange mark the point in splitting where diverging lineages represent unique entities with independent evolutionary trajectories. By identifying the genes that contribute to barriers between recently diverged species we can begin to understand the evolutionary forces that led to their divergence, and ultimately, speciation. This dissertation focuses on two recently diverged species of field cricket, *Gryllus firmus* and *G. pennsylvanicus*, which hybridize in zone stretching from Connecticut to Virginia. We [1] characterize the nature of understudied barriers that occur after mating, but before fertilization, [2] define the role of proteins secreted from the male accessory glands in those barriers, and [3] describe patterns of variation and introgression within the hybrid zone for genes expressed in the male accessory gland. Chapter 1 finds that there are no post-mating prezygotic barriers between *G. pennsylvanicus* females and *G. firmus* males, but that the reciprocal crosses are isolated by barriers that occur after sperm are released from storage, but before fertilization. Chapter 2 finds that seminal fluid proteins secreted from the male accessory glands induce a short-term egg-laying response, but alone cannot explain the normal induction of egg-laying or unsuccessful fertilization in

heterospecific crosses. Chapter 3 lays the foundation for Chapters 4 and 5 by describing patterns of admixture and the influence of environmental variables on species distributions in a previously uncharacterized region of the hybrid zone in Pennsylvania. Chapter 4 compares patterns of introgression in Pennsylvania for genes known to encode seminal fluid proteins and other genes expressed in the male accessory gland. We find no evidence that seminal fluid proteins contribute to barriers operating within the hybrid zone, but identify a number of other candidate barrier genes. Chapter 5 compares the patterns we see in Pennsylvania with an independent transect of the hybrid zone in Connecticut. We find the same genes under selection in both regions of the hybrid zone, suggesting that these genomic regions contribute to maintaining species boundaries.

BIOGRAPHICAL SKETCH

Erica L. Larson (née Davis) was raised in Chewelah, Washington, a very small town in the Northeast corner of the state. For her undergraduate degree she attended Western Washington University in Bellingham, Washington and spent her summers working as a wildland firefighter for the Washington State Department of Natural Resources in northeastern Washington. She majored in Biology with an emphasis in Ecology and spent several years working as a research assistant for Jeannie Gilbert (benthic invertebrate community structure) and Dr. David Hooper (effects of community composition on ecosystem processes). The summer before her senior year she worked as an intern for Dr. Merrill Peterson studying the evolution of male mate choice in a pair of hybridizing leaf beetles. She has been interested in evolutionary process, speciation, sexual selection, and insects ever since. She continued working with Dr. Peterson throughout her senior year, and for two years as a research technician. It was Dr. Peterson who both encouraged her to pursue a graduate degree, and to consider working with Dr. Richard Harrison at Cornell University. From 2006 to 2012 she worked with Dr. Harrison studying the mechanisms and genetic underpinnings of reproductive barriers between hybridizing field crickets. Upon finishing her degree at Cornell University, Erica will move to Missoula, Montana to begin a post-doctorial position at the University of Montana in the laboratory of Dr. Jeffery Good, where she will continue studying speciation genomics and reproductive biology in hybridizing house mice.

To Merrill Peterson,
for your encouragement and support

ACKNOWLEDGMENTS

This work would not have been possible without the support of many people. Rick Harrison has been involved in every aspect of this research. It has been an honor to work with him. His attention to detail, insight and strong opinions have shaped my research and evolution as a scientist. Kerry Shaw has given me a different and valuable point of view that has challenged me to think carefully about my research. She has also given me a greater appreciation for crickets. Brian Lazzaro has been a great teacher and has brought a broader perspective to my committee. Steve Bogdanowicz has been an unofficial committee member from the beginning. I admire his enthusiasm for research and his amazing ability to solve any problem.

The hard work of Jose Andres, Luana Maroja, Greta Hume, Charles Ross, and all those who have worked on the field cricket hybrid zone built the foundation for this research. I am grateful to be a part of such a friendly and supportive research community. I would like to thank my other collaborates, Gui Becker and Tom White, for their contributions to this research. I would also like to thank Linda Cote for helping generate much of these data. Thanks to the undergraduate research assistants who have worked with me in the Harrison lab, particularly Nicholas Salah and Eli Bondra. It has also been wonderful to work with Jeremy Searle organizing the Cornell Evolution Group. His thoughtfulness and kindness have taught me so much about participating in an academic community.

I am grateful to all of the friends that have made Ithaca my home for six years. The members of the Harrison lab have been great friends and colleagues, with particular thanks

to Findley Finseth, Marie Nydam, Ben Hamilton, Luana Maroja, Jose Andres, Jamie Walters, Ryan Thum, Mike Bronksi, Petra Deane, and Ezra Lencer. I am grateful for the denizens of Corson E231: Dave Cerasale, Paulo Llambias, Mike Booth, Morgan Mouchka, Findley Finseth, and Chris Dalton. I will greatly miss our lunchtime conversations and the cynicism many of you brought to my academic perspective. Thanks to other good friends, Dawn Higginson, Katie Wagner, Sarah States and Scott Taylor. Thanks to all of my fellow Whitehousians: Pat Brehl, Mat Langlois, Danica Lombardozzi, Jon DeCoste, Susan Cook, Zach Patton, Marjorie Webber, Chris Bliss, Nancy Chen, Billie Gould, and Ben Dalziel. You have been my family.

I would like to thank the people whose encouragement and support helped bring me to Cornell. Thanks to the folks at Western Washington that got me into research: Jeanie Gilbert, Dave Hooper, Merrill Peterson, Kirsten Monsen and Monique Brewer. Merrill has been a wonderful educator and mentor. I never would have applied to Cornell without his encouragement. My parents, Susan and Darrell Davis, raised me to believe I could do anything. Finally, I owe the greatest thanks to my husband, Jonathan Larson, who has supported me every day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

POST-MATING PREZYGOTIC BARRIERS TO GENE EXCHANGE BETWEEN HYBRIDIZING FIELD CRICKETS

Published as: Erica L. Larson, Gretchen L. Hume, José A. Andrés, and Richard G. Harrison. 2012. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. Journal of Evolutionary Biology 25:174-186.

Abstract

Studies of sexual selection in speciation have traditionally focused on mate preference, with less attention given to traits that act between copulation and fertilization. However, recent work suggests that post-mating prezygotic barriers may play an important role in speciation. Here, we evaluate the role of such barriers in the field crickets, *Gryllus firmus* and *G. pennsylvanicus*. *Gryllus pennsylvanicus* females mated with *G. firmus* males produce viable, fertile offspring, but when housed with both species produce offspring sired primarily by conspecifics. We evaluate patterns of sperm utilization in doubly-mated *G. pennsylvanicus* females and find no evidence for conspecific sperm precedence. The reciprocal cross (*Gryllus firmus* female x *G. pennsylvanicus* male) produces no progeny. Absence of progeny reflects a barrier to fertilization rather than reduced sperm transfer, storage, or motility. We propose a classification scheme for mechanisms underlying post-mating prezygotic barriers similar to that used for premating barriers.

Introduction

Interactions between males and females are mediated by traits that can evolve rapidly via sexual selection, leading to divergence in reproductive systems (mate recognition or fertilization) between isolated populations and potentially resulting in the evolution of reproductive incompatibilities (Lande, 1981; West-Eberhard, 1983). It has long been recognized that closely related, sexually reproducing species exhibit substantial variation in sexual traits (Darwin, 1871; Andersson, 1994), and that these trait differences can act as barriers to gene exchange between species (Dobzhansky, 1937; Mayr, 1963). However, it is only in more recent decades that a

direct role for sexual selection in speciation has gained wider acceptance (Panhuis et al., 2001; Ritchie, 2007; Snook et al., 2009).

Until recently, studies of the role sexual selection in speciation have focused on traits that influence courtship and mate choice. Less attention has been given to traits that act after copulation, but before fertilization, so-called post-mating prezygotic barriers. The cryptic nature of these barriers makes them particularly difficult to detect and study, and although they appear to be widespread, they have been investigated in relatively few systems (reviewed in Markow, 1997; Howard, 1999; Eady, 2001; Howard et al., 2009). Nevertheless, there is strong evidence for the coevolution of male and female post-copulatory reproductive traits (e.g. Rice, 1996; Pitnick et al., 1999; Arnqvist et al., 2000; Pitnick et al., 2003), suggesting that sexual selection can cause divergence in post-mating prezygotic traits as easily as pre-mating traits (Holland & Rice, 1998; Parker & Partridge, 1998; Rice, 1998; Gavrilets, 2000). Similarly, studies across a wide range of animal taxa have shown that reproductive proteins evolve rapidly, and are strikingly divergent among closely related species (reviewed in Civetta & Singh, 1998; Swanson & Vacquier, 2002; Clark et al., 2006). These observations also suggest that post-mating prezygotic barriers may play an important role in speciation.

The interactions between male and female post-copulatory reproductive traits are extraordinarily complex (reviewed in Wolfner, 2009), and, as a result, fertilization in a heterospecific cross can fail at a number of critical steps. Males may transfer fewer sperm, or the sperm may have reduced viability or motility within the reproductive tract of a heterospecific female. Females may store less heterospecific sperm or differentially utilize sperm from storage. Finally, sperm and egg incompatibilities could prevent sperm from locating and binding to the egg, and if the sperm enters the egg, sperm and egg nuclei may not fuse. A consequence of the

complex interactions between the male ejaculate and the female reproductive tract is that small reductions in the efficiency or success of any of these steps could lead to reproductive incompatibilities between species (Coyne & Orr, 2004). Here we examine the potential role of post-mating prezygotic barriers in interactions between the hybridizing field crickets, *Gryllus firmus* and *G. pennsylvanicus*. Multiple pre-mating barriers to gene exchange have already been documented between these recently diverged species (Harrison, 1983, 1985; Harrison & Rand, 1989; Maroja et al., 2009a), but interspecific matings do occur in the field, and what transpires within the reproductive tract of the female remains obscure.

The field cricket system

The field crickets *Gryllus firmus* and *G. pennsylvanicus* interact in a well-characterized hybrid zone stretching from Connecticut to North Carolina, along the eastern edge of the Appalachian Mountains (Harrison & Arnold, 1982). Where the two species co-occur, F1 hybrids are rare, but individuals of mixed ancestry are common (Harrison & Bogdanowicz, 1997; Maroja et al., 2009b). Because female field crickets are promiscuous, and mate with many different males (Solymar & Cade, 1990; Bretman & Tregenza, 2005), there is the potential for post-copulatory sexual selection. During copulation, males transfer to the female a single spermatophore containing sperm and seminal fluid, which the female stores in an elastic spermatheca that expands to hold the ejaculates from multiple males (reviewed in Zuk & Simmons 1997). Seminal fluid proteins contained within the spermatophore have been identified, and some of the genes that encode these proteins are evolving rapidly under positive selection and are strikingly divergent between the two species (Andres et al., 2006; Braswell et al., 2006; Andres et al., 2008; Maroja et al., 2009b).

The two reciprocal crosses between *G. pennsylvanicus* and *G. firmus* have very different outcomes. When *G. pennsylvanicus* females mate with *G. firmus* males, there are no obvious barriers to fertilization, and the offspring produced are both viable and fertile (Harrison, 1983). However, when *G. pennsylvanicus* females are co-housed with a male of each species, most of the offspring are sired by the conspecific male (Harrison & Rand, 1989). This bias in paternity might be driven by *G. pennsylvanicus* females preferentially mating with conspecific males (Maroja et al., 2009a), but post-mating prezygotic barriers (e.g., conspecific sperm precedence) could contribute. Such barriers may not be detected in single-mating crosses between species, where even inefficient fertilization will still result in offspring. However, in double matings, where there is competition with conspecific ejaculates, even minor reproductive incompatibilities can result in substantial conspecific sperm precedence (Howard et al., 1998; Geyer & Palumbi, 2005; Harper & Hart, 2005).

In contrast, the reciprocal cross between *G. firmus* females and *G. pennsylvanicus* males results in few eggs, none of which hatch (Harrison, 1983; Maroja et al., 2009a). The unhatched eggs are indistinguishable from unfertilized eggs deposited by virgin females (Maroja et al., 2008). This suggests that the barrier in this cross occurs prior to fertilization; however, it is also possible that eggs are fertilized but die very early in development. There is no evidence that *Wolbachia* infections have any role in the one-way incompatibility (Maroja et al., 2008). If the eggs are unfertilized, it could be the result of fewer sperm available for fertilization (through reduced transfer and storage of sperm or reduced sperm viability) or a consequence of gametic incompatibility.

Here, we evaluate the role of post-mating prezygotic barriers in limiting gene exchange between these species. We do so in the context of the fundamental asymmetry between the two

reciprocal crosses by 1) evaluating sperm utilization patterns in *G. pennsylvanicus* females mated to both conspecific and heterospecific males and 2) evaluating the nature of barriers that give rise to the one-way incompatibility between *G. firmus* females and *G. pennsylvanicus* males. In the first cross, we find no evidence for differential utilization of conspecific sperm by *G. pennsylvanicus* females, and thus, no evidence for post-mating prezygotic barriers. In the second cross, we find that barriers occur prior to fertilization but post-mating. These, barriers cannot be explained by reduced sperm transfer, storage or motility, and appear to result from an inability of sperm to meet and fuse with the egg. Finally, we suggest that the multiplicity of mechanisms that determine the success of sperm-egg interactions within the reproductive tract of a female deserves a classification scheme similar to that used for premating barriers.

Materials and Methods

Cricket Collections

We collected crickets for sperm storage, motility and precedence experiments in August 1999, 2000, and 2001 from pure populations of *G. firmus* in Guilford, CT (41°16'5"; -72°40'4"); Hammonasset Beach State Park, CT (49°14'59"; -72°32'46"); and Saybrook Point, CT (41°16'51"; -72°20'59") and *G. pennsylvanicus* in Sharon, CT (41°53'0"; -73°29'0") Housatonic Meadows, CT (41°51'44"; -73°22'54"); and Ithaca, NY (42°24'35"; -76°32'46"). In August 2008 and 2009 we collected crickets from two of these populations, Guilford, CT and Ithaca, NY for fertilization experiments. All crickets were collected as late instar nymphs and maintained in plastic cages (30x16x19cm) with food (cat food and rabbit food), a water vial, and egg flats for shelter. The cages were maintained under a 12:12 hour light/dark cycle at 25°C.

Matings

For all crosses, virgin females six to ten days post-eclosion and virgin males six to twenty-one days post-eclosion were randomly assigned to treatments (see below). For each experiment we label treatments to indicate species (F = *G. firmus*, P = *G. pennsylvanicus*) and the order of matings, with the first letter indicating the female, and the subsequent letters indicating the males with which she mated (e.g., PFF represents a *G. pennsylvanicus* female mated in succession to two *G. firmus* males). Females and males were placed in a petri dish (9 cm) lined with moistened filter paper to provide traction. In *Gryllus*, males produce and hold a spermatophore in the genital tract prior to encountering a female and thus are unable to adjust their ejaculate expenditure in response to individual mates. We observed all mating trials to ensure that the spermatophore was properly attached. Previous studies of crickets have demonstrated that spermatophore attachment time is extended by male postcopulatory mate guarding (Khalifa 1950; Simmons 1986; Evans 1988; Sakaluk 1991), and the mean duration of guarding appears to correspond to the time required for complete sperm transfer (Simmons 1986). To standardize the spermatophore attachment time, we left mated pairs undisturbed, allowing the males to guard the females until they reinitiated courtship (approximately 40 minutes), after which, the females were isolated. For double matings, females were mated using the exact same procedure 24 hours following the first mating. Pairs that did not successfully transfer a spermatophore within an hour were excluded from experiments.

Sperm competition

Gryllus pennsylvanicus males were randomly assigned to one of four treatments. For each treatment, females were mated to a first male and then mated to a second male after 24

hours. In treatments PPP and PFF females mate twice with either two different conspecific males (N_{PPP} = 28 total, 19 successful) or heterospecific males (N_{PFF} = 27 total, 20 successful) and in treatments PPF and PFP females were mated twice either first with a conspecific male, then a heterospecific male (N_{PPF} = 31 total, 22 successful) or first a heterospecific male, then a conspecific male (N_{PFP} = 29 total, 25 successful).

Following matings, females were isolated in individual plastic cages and provided with a petri dish (9 cm) filled with a mixture of moistened soil and sand for an oviposition substrate. The petri dishes were replaced every ten days, and incubated at 25^{o}C for 21 days and then at 10^{o}C for 102 days to ensure synchronous hatching (Harrison, 1985). For hatching, eggs were removed from chill and incubated at 25^{o}C. Females that produced less than ten progeny were excluded from analysis. For each female we collected up to 20 first instar nymphs and stored them at -80^{o}C until paternity analysis.

For paternity analysis, genomic DNA was extracted using DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA, USA) from an adult femur or an entire early instar nymph. Microsatellite markers were developed from *G. pennsylvanicus* genomic DNA using the methods outlined in Hamilton et al. (1999). DNA was digested with the restriction enzymes ApoI and BstyI (New England Biolabs, Ipswich, MA), and DNA fragments ranging from 600 to 900bp were then ligated to pUC 19 plasmids (previously digested with EcoRI and BamHI). These ligations were used to transform *E. coli* DH5-α cells and were plated on Luria-Bertani (LB) agar/ampicillin plates. Colonies were transferred to nylon membranes that were then hybridized to ^{33}P radiolabeled probes composed of di- and trinucleotide probes [GT8, TC9.5, TTA12, GAT7, GTT6.33, GTA8.33, TTC7, GCT4.33, GTG4.67, GTC4.67, TCC5]. Inserts (149 clones) were amplified using M13 primers and sequenced on an ABI 3100 DNA sequencer (Applied

Biosystems Inc., Foster City, CA). Sequences were trimmed and aligned using SEQMAN PRO (DNASTAR, Inc., Madison, WI). We designed primers for unique sequences with the program PRIMERSELECT (DNASTAR, Inc., Madison, WI).

These locus-specific primers were used to polymerase chain reaction (PCR) amplify microsatellite loci in 10μL reactions containing 1μL genomic DNA, 2mM MgCl2, 0.2μM of each forward and reverse primer and 0.1μL (0.5U) Platinum Taq polymerase (Invitrogen, San Diego, CA) in 1x PCR buffer (20mM Tris-HCL, pH 8.4, 50mM KCl). PCRs for all primer pairs were performed using an initial denaturation at 95^{o}C for two minutes followed by 35 cycles of 95^{o}C for 50 seconds, primer specific annealing temperate for one minute and 72^{o}C for one minute. The forward primer was labeled with a fluorescent tag (6-FAM, PET, NED, or VIC). Fluorescent PCR products were diluted 1:15 in water and mixed with formamide and Genescan LIZ-500 size standard (Applied Biosystems) and run on an ABI PRISM 3100 DNA analyzer. Alleles were called using GENEMAPPER (Applied Biosystems) and then verified by eye.

We quantified genetic variation in one population of each species (*G. pennsylvanicus*: Ithaca, NY; *G. firmus*: Guilford, CT) (Table 1.1). We converted our data from a GENEMAPPER (Applied Biosystems) file format using MICROSATELLITE ANALYZER v. 4.05 (Dieringer & Schöltterer 2003). Tests for linkage disequilibrium and deviation from Hardy-Weinberg equilibrium were performed using GENEPOP v. 4.1 (Raymond & Rousset 1995) and we used CERVUS v. 3.0 (Kalinowski et al. 2007) to test parentage exclusion probabilities, estimate null alleles and the polymorphic information content of the markers.

To genotype parents and offspring from our crosses we used markers 14, 143 and PGI and to genotype individuals that were ambiguous we used a combination of the remaining five markers. The high level of polymorphism in our genetic markers allowed us to assign paternity

Table 1.1 Primer sequences and amplification conditions for eight *Gryllus* microsatellite loci used in paternity analysis. T_a, annealing temperature (^oC); Size, allele size range in base pairs; Sp, species; N_a, number of alleles; N, number of individuals scored; H_O, observed heterozygosity; H_E, expected heterozygosity; PIC, polymorphic information content; Null, frequency of null alleles. Numbers in bold indicate significant deviations from Hardy-Weinberg equilibrium; *represent significant deviations following Bonferroni correction.

Locus	Repeat	Primer sequence (5′–3′)	T_a	Size	Sp	N_a	N	H_O	H_E	PIC	Null	GenBank
Gr116	$A(T)C_{59}$	CACACGTGATATACCGACAACACT	61	186–302	*Gp*	35	44	**0.841**	**0.972**	0.960	0.067	JN375327
		GCTTTAACCCCCTAGCAGAATAAT			*Gf*	27	26	**0.692**	**0.970***	0.970	0.161	
PGI	ATT_{15}	GAATGCATACATCAGTGTCATGAACA	56	181–520	*Gp*	30	42	0.881	0.959	0.945	0.035	JN379460
		TGACTCAAAATAAGCATTATTTCAGC			*Gf*	30	27	0.963	0.970	0.949	0.006	
Gr14	ATT_{17}	AGAAACGCTAGGTTAGTACTATTGGTAA	56	148–283	*Gp*	35	46	0.913	0.971	0.959	0.026	JN375323
		TCGCGAAATTCGAATTGTATC			*Gf*	27	28	0.964	0.971	0.951	0.006	
Gr143	TG_{11}	CTGCCGCATTCACCAATCATTCAACTAT	58	152–224	*Gp*	29	46	**0.870**	**0.952**	0.939	0.041	JN375328
		CAACCAAGGGGCAAAATGAGTCAAACTT			*Gf*	18	28	0.893	0.923	0.899	0.008	
Gr54	AC_{23}	CGCGCCATCTTTTTCTTATTCATTCTTCT	62	157–204	*Gp*	28	45	**0.578**	**0.953***	0.939	0.242	JN375324
		CGTTCCTCTCCGCTGCCGTCGTA			*Gf*	18	27	**0.778**	**0.943**	0.920	0.085	
Gr112	CT_{25}	GAATAATAAAAGCGCAAGGATA	57	204–368	*Gp*	42	46	**0.630**	**0.951***	0.938	0.202	JN375325
		CACCATTACACGTACTCACTCG			*Gf*	26	27	**0.481**	**0.913***	0.891	0.305	
Gr115a	TAA_{21}	TCGCCAAGACTTACGCTCAG	61	190–389	*Gp*	24	23	**0.565**	**0.966***	0.942	0.253	JN375326
		CCTCCTCGGCACAGACG			*Gf*	30	26	**0.667**	**0.973***	0.952	0.161	
Gr145	AAT_{42}	AACACTTGGCAACTTTGATTCT	58	117–301	*Gp*	17	20	**0.200**	**0.944***	0.915	0.644	JN375329
		TCAGTGAGGCTTAGTTTGTGT			*Gf*	19	22	**0.090**	**0.957***	0.930	0.823	

by eye. To estimate sperm precedence, we calculated the observed proportion of offspring sired by the second male (P_2). The extent of sperm precedence was analyzed using a generalized linear model (GLM) with binomially distributed error and a logistic link function (Crawley 1993; Wilson & Hardy 2002). Analyses were performed using JMP v. 8.0.1 (SAS Institute Inc. Cary, NC).

Sperm storage and motility

To evaluate the importance of sperm transfer and function in the incompatible cross, we used estimates of sperm storage and motility. In *Gryllus*, severe sperm clumping occurs within the female reproductive tract; therefore it is not practical to score individual sperm presence or motility. Similarly, estimates of viability using Live/Dead^{TM} sperm staining kits do not provide reliable quantitative results. To estimate sperm transfer and storage, we first tested if there was a correlation between spermatheca volume and number of matings, which approximates sperm transfer (Simmons 1986). We consecutively mated *G. firmus* females to one, two or more than three *G. firmus* males and dissected females 24 hours following the last mating. Prior to dissection, females were anesthetized by chilling at 4^{o}C for one hour. Spermatheca were removed and placed in a drop of phosphate buffered saline (PBS). We measured the spermatheca volume calculated as the volume of an ellipsoid using the formula $(4/3)\pi abc$, where a and b are the equatorial radii and c is the polar radius. All measurements were made at 50x magnification.

To estimate differences in sperm storage and motility between conspecific and heterospecific matings we randomly assigned *G. firmus* females to mate with either a *G. firmus* male (N_{FF} = 13 total, 11 successful) or a *G. pennsylvanicus* male (N_{FP} = 18 total, 17 successful) and then dissected females 24 hours after the second mating. We dissected females, and

measured the spermatheca volume using the methods described above. We then gently opened the spermatheca with dissecting pins, and at 400x magnification estimated the percent of visibly motile sperm at three points on the slide. We tested for a correlation between spermatheca volume and number of matings using a linear regression and compared mean spermatheca volume and percent motile sperm of female *G. firmus* mated to conspecifics with *G. firmus* females mated to heterospecifics using the t-test implemented in JMP v. 8.0.1.

Fertilization

To determine whether unhatched eggs from *G. firmus* females mated with *G. pennsylvanicus* males are unfertilized or if embryos die early in development, we visualized egg nuclei using the cell permeable stain, 4',6-diami-2-phenylindole (DAPI). In *Gryllus*, sperm enter the egg while the egg is in the genital chamber just prior to oviposition (Sugawara & Loher, 1986). Immediately following oviposition, a single nucleus (primary oocyte nucleus) is visible just below the dorsal surface of the egg. Within one hour the nucleus divides into two nuclei (secondary oocyte nucleus and polar body), and within three hours there are three or four nuclei (female pronucleus and 2-3 polar bodies) visible. After the completion of meiotic division, the female pronucleus migrates to the ventral side of the egg where male and female pronuclei fuse, and mitotic division begins (Sato & Tanaka-Sato, 2002). Unfertilized eggs remain arrested in metaphase of the first meiotic division unless activated by an external signal, presumably the absorption of water from the oviposition substrate (Sarashina et al., 2003). Once activated, unfertilized eggs may continue through meiotic division, but progression to mitosis has never been observed. Because of the difficulty of visualizing nuclei with DAPI during very early stages of development, we divided stained eggs into two categories (Matute & Coyne, 2010).

We considered eggs fertilized if nuclei resulting from mitotic division were observed (≥4 nuclei on the ventral side of the egg or dispersed throughout the egg cytoplasm). Similarly, we considered eggs unfertilized if we only observed nuclei resulting from meiotic division (≤4 nuclei on the dorsal side of the egg) or no nuclei.

We evaluated fertilization for eggs laid by *G. firmus* females which were virgin (N_F = 10), mated with *G. firmus* males (N_{FF} = 41) or *G. pennsylvanicus* males (N_{FP} = 47). To ensure that females had no shortage of sperm, each mating pair was observed until the transfer of the first spermatophore, and then co-housed for an additional 24 hours to allow for multiple matings. Females were then separated into groups and provided with oviposition substrate. To evaluate fertilization we left the oviposition substrate for 24 hours and then incubated eggs for an additional 24 hours at 25°C. To evaluate sperm entry into the egg, we left the oviposition substrate for five hours, and then processed the eggs immediately.

We stained eggs with DAPI using methods adapted from Sarashina et al. (2003, 2005). Eggs were collected in glass scintillation vials with a paintbrush, dechorionated with 50% bleach for five minutes with gentle pipetting, and then rinsed thoroughly with PBS. Eggs were fixed in a solution of equal parts heptane and paraformaldehyde with gentle shaking at 25°C for 20 minutes followed by devitellinization with methanol and storage in methanol at 4°C. To visualize nuclei, we stained eggs with 0.2ug/mL DAPI in PBS for 20 minutes with gentle shaking at 25°C. The eggs were examined under a fluorescence microscope.

To confirm that patterns of development (position of meiotic nuclei and polar bodies) in unfertilized *G. firmus* eggs correspond to our expectations, we visually compared the distribution of nuclei in eggs laid by virgin *G. firmus* females (F) with patterns observed in *G. bimaculatus* (Sato and Tanaka-Sato 2002; Sarashina et al. 2003; Sarashina et al. 2005). We then compared

the proportion of fertilized eggs laid by *G. firmus* females that were mated with either conspecifics (FF) or heterospecifics (FP) at both five hours and 48 hours after oviposition, using the Fisher's exact test implemented in JMP v. 8.0.1.

Results

Sperm competition

Six microsatellite loci in *G. pennsylvanicus* and four microsatellite loci in *G. firmus* deviated from Hardy-Weinberg equilibrium after Bonferroni correction (Table 1.1), most likely as a result of the presence of null alleles. There was no evidence of linkage disequilibrium between any pair of loci. Despite the presence of null alleles, the combined non-exclusion probability of the second parent across all eight loci was 2.0×10^{-8} and 6.0×10^{-8} for *G. pennsylvanicus* and *G. firmus* respectively, indicating these markers are suitable for assigning paternity. Of the 1,342 nymphs collected from doubly-mated *G. pennsylvanicus* females, 95.8% were genotyped and assigned paternity successfully. The remaining 4.2% had poor microsatellite amplification either due to DNA degradation, impurities in the DNA extraction, or the presence of null alleles. No non-parental alleles were observed in the offspring. There were no differences between treatments in the proportion of offspring sired by the second male (Figure 1.1, GLM: $X^2_{3\text{d.f.}} = 0.483$, $P = 0.9226$). Thus, there was no evidence for either first or second male sperm precedence within species (PPP and PFF crosses) or conspecific sperm precedence in interspecific crosses (PFP and PPF).

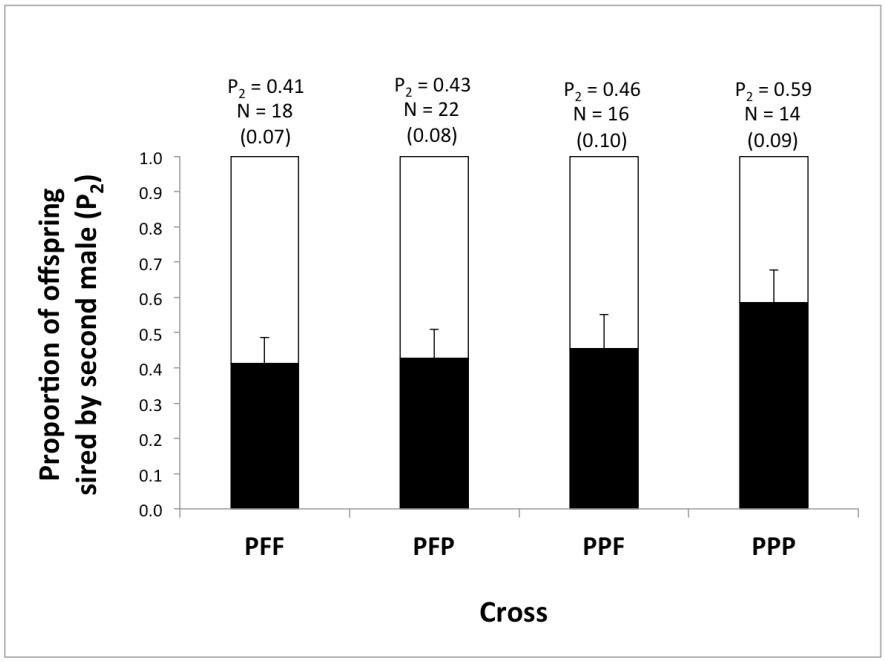


Figure 1.1 Mean (SE) proportion of offspring per *Gryllus pennsylvanicus* female sired by first (open) and second (solid) male. Mean P_2-values (SE) are shown above bars (N = sample size).

Sperm Storage and Motility

Gryllus firmus spermatheca volume increased significantly with numbers of matings (Figure 1.2, $R^2 = 0.42$, $F_{1d.f.} = 16.09$, $P = 0.0006$). *Gryllus firmus* females mated with *G. pennsylvanicus* males stored the same volume of sperm as *G. firmus* females mated with conspecific males (Table 1.2, t-test: $t_{22.5d.f.} = 1.38$, $P = 0.182$). Similarly, motile sperm were observed in the spermatheca of females from every cross, and the percent of motile sperm did not differ between those mated with conspecific males and heterospecific males (Table 1.2, t-test: $t_{7.8d.f.} = 2.16$, $P = 0.063$).

Fertilization

In eggs from virgin *G. firmus* females (N = 57) we saw the same stages of meiotic division that have been observed in *G. bimaculatus* (always ≤4 nuclei positioned on the dorsal side of the egg). We used this as our as reference for categorizing eggs as fertilized or unfertilized (see Methods). We found a significant difference in the proportion of fertilized eggs between treatments (Table 1.2, Fisher's exact test: $X^2_{1d.f.} = 432.60$, $N = 426$, $P = <0.0001$). The majority of the eggs from conspecific crosses (FF = 96.86%) were fertilized, and development proceeded normally (Figure 1.3B). In contrast, most eggs from heterospecific crosses were unfertilized (FP = 95.88%) (Figure 1.3C). The small percentage of heterospecific eggs that were fertilized underwent mitotic division, but died early in development (Figure 1.3D). Egg meiotic nuclei were observed in nearly half (FP = 48.8%) of all unfertilized eggs, but sperm nuclei were never observed. Our evaluations of sperm entry into the egg were limited by our ability to collect eggs immediately after oviposition. Females were reluctant to oviposit when disturbed frequently for egg collections. In eggs from conspecific crosses, sperm nuclei were visible in

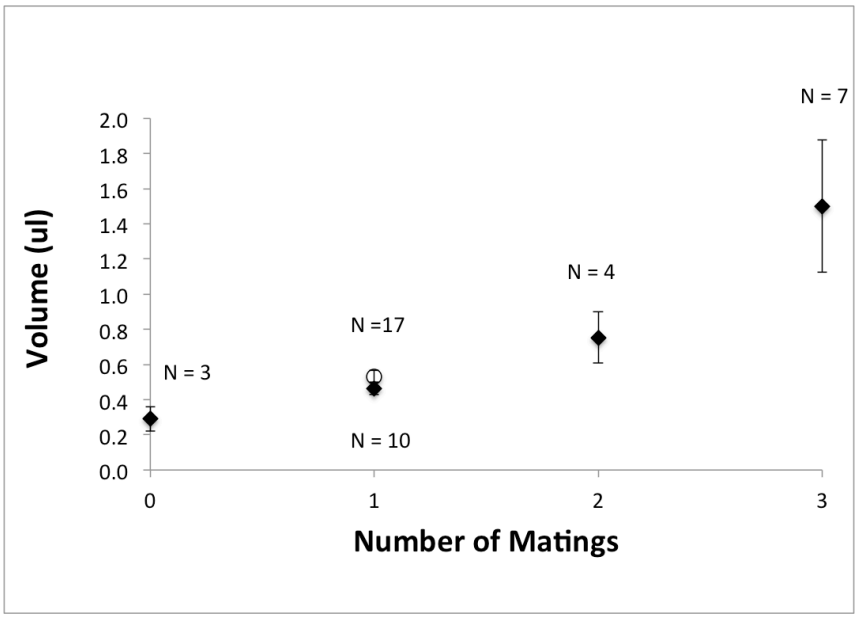


Figure 1.2 Relationship between mean (SE) spermatheca volume of *Gryllus firmus* females and number of matings with a *G. firmus* male (solid) (N = sample size). For a comparison, the spermatheca volume of *G. firmus* females mated with a single *G. pennsylvanicus* male (open) is overlaid

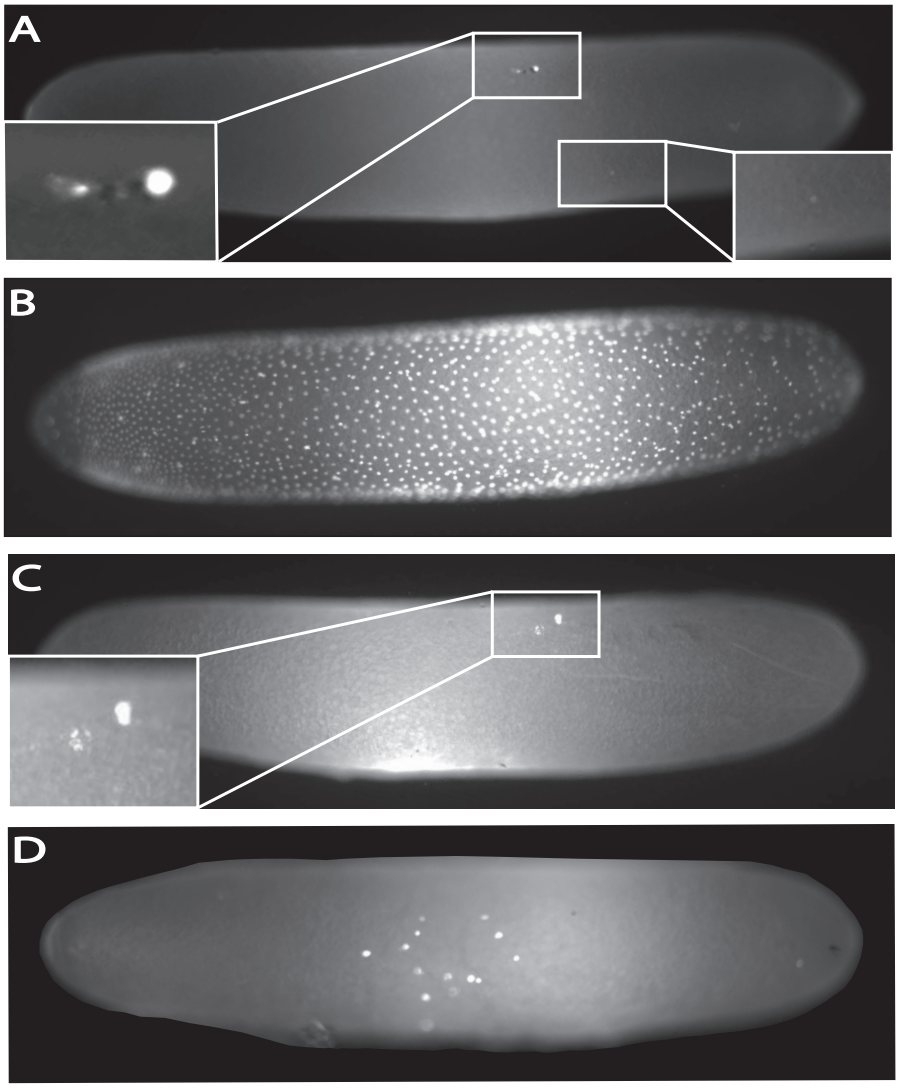


Figure 1.3 Distribution of nuclei stained with DAPI (white spots) during early development of *G. firmus* eggs. The anterior of the egg is to the left and dorsal is the top. A) An egg from a *G. firmus* female mated with a conspecific male 0-5 hours after oviposition. Two nuclei (female pronucleus and 1 polar body) are visible just below the dorsal surface and single nucleus (male pronucleus) is visible near the ventral surface. B) An egg from a *G. firmus* female mated with a conspecific male 24-48 hours after oviposition. The egg has approximately 2000 nuclei, most of which are distributed at the periphery of the egg. C) An egg from a *G. firmus* female mated with a heterospecific male 24-48 hours after oviposition. Two nuclei (female pronucleus and 1 polar body) are visible just under the dorsal surface. D) An egg from a *G. firmus* female mated with a heterospecific male 24-48 hours after oviposition. The egg has approximately 14 nuclei that have not reached the periphery of the egg.

Table 1.2 Spermatheca volumes and sperm viability in *G. firmus* females mated with either *G. firmus* males or *G. pennsylvanicus* males. See Methods for details. Embryo staining estimates (1) the proportion of eggs that are fertilized, using eggs harvested 48 hours after the beginning of oviposition, and (2) the proportion of eggs with sperm nuclei visible after 5 hours of development.

		G. firmus		*G. pennsylvanicus*		
	Trait	Value	*N*	Value	*N*	*P*-value
Spermatheca dissections	Sperm storage	0.46 µL (0.04)	10	0.53 µL (0.03)	17	0.2004
	Sperm motility	80% (2.0)	7	85% (0.9)	10	0.0593
Embryo staining	Fertilization (48 h)	96.86%	159	4.12%	267	< 0.0001
	Sperm visible (5 h)	63.16%	19	0%	7	0.0064

over half the eggs, (FF = 63.16%; see Figure 1.3A) whereas in heterospecific crosses sperm nuclei were never observed (Table 1.2, Fisher's exact test: $X^2_{1d.f.} = 10.88$, $N = 26$, $P = 0.0064$).

Discussion

Previous studies of the field cricket hybrid zone have characterized pre-mating barriers that operate to limit or prevent gene exchange. These include temporal isolation and habitat or ecogeographic isolation. Additional barriers are also present, but when and how they act has remained unclear. The observation that female *G. pennsylvanicus* exposed to males of both species (either in the lab or in mixed populations in the field) appear to produce offspring sired predominantly by conspecific males could be explained by assortative mating and/or conspecific sperm precedence. The one-way incompatibility between *G. firmus* females and *G. pennsylvanicus* males could result from failures at a number of steps in the transfer or utilization of sperm or in the ability of sperm to fertilize the egg. Here we have provided data that clarify the nature of these barriers.

No sperm precedence in G. pennsylvanicus females

When the ejaculates of multiple males compete within a female reproductive tract for fertilization, it is common for paternity to be biased in favor of one or the other male (Parker, 1970). Differences in sperm utilization can occur through a reduction in sperm numbers or sperm precedence. Sperm can be reduced in storage either passively through loss between matings or usage, or they can be actively lost through displacement by competing males or ejection by the female (Eberhard 1996). Similarly, sperm precedence can occur through passive mechanisms, such as sperm stratification, or active mechanisms, such as sperm competition or

sperm selection by the female (reviewed in Simmons, 2001). Given that there is substantial intraspecific variation in sperm competitive ability, interspecific sperm competition is likely to be important, and may lead to widespread post-mating prezygotic barriers among hybridizing taxa (reviewed in Howard, 1999; Howard et al., 2009).

We find no evidence of differential fertilization in *G. pennsylvanicus* females mated with two conspecific males (or two heterospecific males). In both cases, the second male sires half the offspring when conspecific males compete for fertilization. This is contrary to what is found in many insects, which often have high values of P_2 (Simmons & Siva-Jothy, 1998; Simmons, 2001). In Orthopterans, patterns of paternity bias are predominantly due to the relative number of sperm stored from different males at the time of fertilization (Table 1.3). In several species, there are mechanisms of active sperm displacement (Ono et al., 1989; Helversen & Helversen, 1991), but more typically there is variation in the amount of sperm transferred (Sakaluk, 1986; Gwynne & Snedden, 1995; Sakaluk & Eggert, 1996), or passive sperm loss (Parker & Smith, 1975; López-León et al., 1993; Reinhardt, 2000, Reinhardt & Meister, 2000; Zhu & Tanaka, 2002; Hockham et al. 2004). However, intermediate values of P_2 are observed in most crickets (Backus & Cade, 1986; Simmons, 1986; Gregory & Howard, 1994; Morrow & Gage, 2001, Bussière et al., 2006; Bretman et al., 2009; Hall et al., 2010). Crickets have a round, elastic spermatheca, that expands to hold the ejaculates of multiple males, which likely promotes sperm mixing. Observed variance in patterns of paternity bias is typically due to mechanisms of sperm competition avoidance, such as mate guarding, and sperm loading through repeated copulations. In addition, females can exert control over male paternity by early removal of the externally attached spermatophore (Sakaluk, 1984; Simmons, 1986, 1987; Sakaluk & Eggert, 1996; Bussière et al., 2006). Consistent with this, we found intermediate values of P_2 when sperm from

Table 1.3 Summary of intraspecific sperm utilization patterns in Orthopterans, expressed either as the proportion of offspring fathered by the second male (P_2) or the proportion of sperm from the second male in the female storage organ (S_2). Values for sperm utilization are range and standard deviation (SD). Updated from Simmons and Siva-Jothy (1998) and Simmons (2001).

Taxa	Mean P_2/S_2	Range	SD	Explanation	Reference
Gryllidae					
Gryllus bimaculatus	0.33/0.45/0.68*	0.00–0.80	0.06/0.03/0.06	Spermatophore	Simmons (1987)
	0.45/0.56†	–	0.4/0.36		Morrow & Gage (2001)
	0.38/0.84‡	0–1	0.07/0.04	Relatedness attachment	Bretman *et al.* (2009)
Gryllus integer	0.72	0.08–1.00	–		Backus & Cade (1986)
Gryllodes sigillatus	0.42	0.04–0.88	0.20	Spermatophore attachment	Sakaluk (1986), Sakaluk & Eggert (1996)
Teleogryllus commodus	NA	0.25–0.55	0.04–0.07	Spermatophore attachment	Bussière *et al.* (2006), Hall *et al.* (2010)
Telogryllus oceanicus	0.46	0.05–0.86	0.24		Simmons *et al.* (2003)
Allenombius fasciatus	0.62/0.02/1.0§	–	–	PMPZ	Gregory & Howard (1994)
Allenombius socius	0.43/0.01/0.95§	–	–	PMPZ	Gregory & Howard (1994)
Truljalia hibinonis	0.88	–	0.12		Ono *et al.* (1989)
Tettigoniidae					
Decticus verrucivorus	0.50	0.03–1	0.31	Spermatophore size	Wedell (1991)
Requena verticalis	0.00–0.19	0–0.94	0.17	Spermatophore size; re-mating interval	Gwynne & Snedden (1995)
Kawanaphila nartee	0.69	0.13–1	0.33		Simmons (1995)
Poecilimon veluchianus	0.90	0.87–0.93	0.03		Achmann *et al.* (1992)
Metaplastes ornatus	0.85	–	–		Helversen & Helversen (1991)
Ephippiger ephippiger	0.75	0–1	0.43	Re-mating interval	Hockham *et al.* (2004)
Steropleurus stali	0.95	0.08–1.00			Vahed (1998)
Acrididae					
Schistocerca gregaria	1.00	–	–		Hunter-Jones (1960)
Locusta migratoria	0.38–0.86	0–1	–	Re-mating interval, sperm displacement	Parker & Smith (1975)
	0.61	0.06–1	0.32	Copulation duration, sperm ejection	Gregory (1965), Reinhardt & Meister (2000), Zhu & Tanaka (2002)
Paratettix texanus	0.58	0–0.91	0.37		Nabours (1927)
Podisma pedestris, XO	0.55/0.42¶	0.29–1/0–1	0.10, 0.14		Hewitt *et al.* (1989)
Podisma pedestris, XY	0.54/017¶	0.08–1/0–0.31	0.20, 0.06	PMPZ	Hewitt *et al.* (1989)
Eyprepocnemis plorans	0.92	0.40–1	–	Times mated	López-León *et al.* (1993)
Chorthippus p. erythropus	0.64/0.57¶	–	–		Bella *et al.* (1992)
Chorthippus p. parallelus	0.84/0.35¶	–	–	PMPZ	Bella *et al.* (1992), Reinhardt (2000)
Chorthippus biguttulus	~0.92	0.77–1	0.13		Reinhardt (2000)

PMPZ, post-mating prezygotic barriers, *females mated once, twice and thrice; †males breed for sperm with short tails and long tails; ‡mated pairs are siblings and unrelated; §females mated with two conspecific males, a heterospecific followed by a conspecific male and a conspecific followed by a heterospecific male; ¶females mated with a heterospecific male followed by a conspecific male and a conspecific male followed by a heterospecific male.

two males of the same species compete for fertilization. Because we did not allow for sperm competition avoidance mechanisms other than spermatophore removal, we saw very little deviation from the mean.

Intermediate intraspecific values of P_2 do not necessarily mean that sperm precedence is unlikely when the ejaculates of conspecific and heterospecific males compete. With the ejaculates of the first and second male mixed within the female reproductive tract, sperm competition is high, and heterospecific sperm may be at a disadvantage. Conspecific sperm precedence has been observed in both ground crickets and grasshoppers (Hewitt et al., 1989; Bella et al., 1992; Gregory & Howard, 1994). However, in *G. pennsylvanicus* females, heterospecific and conspecific sperm are equally likely to fertilize, regardless of mate order. Our results suggest that the sperm of both species are equally competitive within the *G. pennsylvanicus* reproductive tract and that there is no cryptic choice by *G. pennsylvanicus* females. This study is one of the few to document the absence of conspecific sperm precedence between closely related taxa (reviewed in Howard et al. 2009).

The absence of sperm precedence in doubly-mated *G. pennsylvanicus* females means that assortative mating (behavioral isolation) must explain previous observations that *G. pennsylvanicus* females, given the opportunity to mate with males of both species, produce offspring mostly sired by conspecific males (Harrison, 1986; Harrison & Rand, 1989). The basis for mate choice in this system is not yet clear, although song is not a likely candidate. The calling songs of the two species are slightly different (with substantial overlap) (Doherty & Storz, 1992), but the courtship songs are nearly identical, and "choice" in a lab context occurs in the absence of calling song (Larson, personal observation). Recent behavioral observations indicate that *G. pennsylvanicus* females refuse to mate with heterospecific males more often than with

conspecific males (Maroja et al., 2009a), suggesting that *G. pennsylvanicus* females are discriminating based on a non-acoustic signaling or behavioral cues during courtship. In particular, cuticular hydrocarbons may mediate close-range courtship and mate choice as has been observed in other crickets (e.g. Tregenza & Wedell, 1997; Mullen et al., 2007; Thomas & Simmons, 2010).

What explains the one-way incompatibility between G. firmus females and G. pennsylvanicus males?

We have shown that there are no obvious differences in sperm transfer or storage between *G. firmus* males and *G. pennsylvanicus* males, and that conspecific and heterospecific sperm appear equally motile within the female *G. firmus* reproductive tract. Thus, the one-way incompatibility must either occur after fertilization or involve a failure to utilize sperm from storage or in the ability of sperm to fertilize the egg. Several lines of evidence have suggested that the incompatibility between *G. firmus* females and *G. pennsylvanicus* males is prezygotic. The eggs from these crosses closely resemble eggs laid by virgin females in both color and shape (Maroja et al., 2008), which can be considered a sign that eggs are unfertilized (Lorch & Servedio 2005; Sweigart 2010). Here, we have confirmed that the majority of the eggs from these heterospecific crosses are unfertilized.

The few eggs from the heterospecific cross in which mitotic division was observed, are possibly a result of successful fertilization and early embryonic mortality. Differences in the fertilization success suggest there could be genetic variability for the trait(s) underlying the one-way incompatibility, and even the potential for inter-population differences across the hybrid zone as is seen in *Drosophila* (see Kelleher & Markow, 2007). Alternatively, the few cases of

mitotic division could be the result of activation of an unfertilized egg. In insects, a range of developmental stages has been observed in unfertilized eggs that have been activated in vitro (reviewed in Went, 1982). It is believed this flexibility in development is the result of gonomeric fertilization in which the female and male pronuclei form independent mitotic spindles and do not fuse until the first cleavage (Kawamura, 2001; Sato & Tanaka-Sato, 2002,). In some insects, the female pronucleus has the ability to continue mitotic divisions in the absence of male chromosomes. Indeed, development and hatching of eggs deposited by old virgin *Gryllus* females has been reported (see Harrison 1983).

Sperm nuclei were never observed in any eggs from the heterospecific cross, while the egg meiotic nuclei were observed in nearly half the eggs at later developmental stages. We also confirmed that in early stages of development, sperm nuclei are visible in eggs fertilized by conspecifics, but are never visible in eggs deposited after mating with heterospecifics. Because we have no *a priori* reason to expect sperm nuclei to degrade at a different rate from meiotic nuclei, our survey of heterospecific eggs from both early and late stages of development is consistent with the absence of sperm inside the egg except in the few instances of apparent successful fertilization.

Given that heterospecific sperm do not appear to enter the egg, the one-way incompatibility is due either to the differential use of sperm from within the spermatheca, the inability of heterospecific sperm to locate the egg within the female reproductive tract, or sperm and egg incompatibilities. Further studies will be needed to identify the precise timing and mechanisms of this barrier. Similar patterns of viable sperm in the sperm storage organ, yet failure for eggs to be fertilized have been observed in at least one other species pair, *Drosophila virilis* and *D. americana* (Sweigart, 2010). The failure of *G. pennsylvanicus* sperm to

successfully fertilize *G. firmus* females could be the result of divergence in male reproductive proteins that mediate these steps or in the interaction of these proteins with the female reproductive tract. Comparisons between *G. pennsylvanicus* and *G. firmus* provide evidence that cricket reproductive proteins diverge rapidly under positive selection (Andres et al., 2006; Braswell et al., 2006; Andres et al., 2008; Maroja et al., 2009b).

Post-mating prezygotic barriers are diverse

Although *G. pennsylvanicus* and *G. firmus* are closely related, and show little genetic divergence (Harrison & Arnold, 1982; Harrison & Bogdanowicz, 1997; Broughton & Harrison, 2003; Andres et al., 2008), these crickets appear to be separated by multiple barriers to gene exchange and F1 hybrids are rare in natural populations (Harrison, 1986; Harrison & Bogdanowicz, 1997). Premating barriers may substantially reduce gene exchange, but opportunities for interspecific matings exist in mixed populations within the hybrid zone, and the observed failure of crosses between *G. firmus* females and *G. pennsylvanicus* males may be an important post-mating barrier.

Coyne and Orr (2004) distinguish between competitive and non-competitive forms of post-mating prezygotic isolation. However, many of the underlying mechanisms of these two forms of isolation are effectively the same; the difference lies in our ability to detect the barriers (Howard et al., 2009). For example, if a male's sperm has reduced motility in the reproductive tract of a heterospecific female, we would be unlikely to notice this trait in a single mating, in which even sperm with low motility could eventually fertilize an egg. But when heterospecific sperm compete for fertilization with conspecific sperm, this same trait could lead to complete reproductive isolation. Although there are no doubt exceptions, in general traits that reduce the

ability of heterospecific sperm to fertilize an egg can either prevent fertilization entirely, and thus be detected in single matings, or these same traits may simply slow the process of fertilization and thus be observed only when there is competition between two or more ejaculates.

Furthermore, classifying post-mating prezygotic barriers as either competitive or non-competitive fails to capture the diversity of mechanisms that mediate post-mating prezygotic barriers. We suggest that such barriers might appropriately be classified using the framework outlined by Mayr (1963) for pre-mating and postzygotic barriers (see Table 1.4). That is, during copulation and within the female reproductive tract, we can recognize barriers that prevent sperm and egg from meeting (barriers involving sperm transfer, sperm viability, sperm utilization), barriers that prevent sperm and egg from fusing if they meet (various sorts of gametic incompatibility), and barriers that operate after sperm-egg fusion (intracellular gametic incompatibility). Understanding the mechanistic basis of such barriers will no doubt be facilitated by ongoing research on the nature and function of rapidly evolving seminal fluid proteins, on the details of fertilization biology, and on comparative aspects of the sperm proteome.

Acknowledgements

We thank S.M. Bogdanowicz for assistance developing microsatellite markers, J.D. Larson for help with cricket collections, and C. Bayles for technical assistance with microscopy. All microscopy was done at the Cornell University's Life Sciences Core Laboratories Center Microscopy and Imaging Facility. This work was supported by a National Science Foundation grant DEB-0639904 to RGH and an AAUW American Fellowship to ELL.

Table 1.4. Summary of barriers to gene exchange between the hybridizing field crickets, *Gryllus firmus* and *G. pennsylvanicus.* Updated from Maroja et al. (2009)

Classification	Barrier	*G. firmus* and *G. pennsylvanicus*	Reference
Premating			
Prevents meeting	Habitat	Association with different soil types	Rand & Harrison (1989), Ross & Harrison (2002, 2006)
	Temporal	Differences in time of adult appearance (due to differences in development time)	Harrison (1985)
	Behavioural	Differences in calling song	Alexander (1957), Doherty & Storz (1992)
Prevents mating	Behavioural	*Gryllus pennsylvanicus* females 'choose' to mate more often with conspecific males	Harrison & Rand (1989), Maroja *et al.* (2009a)
		Longer copulation latency in *G. firmus* females paired with *G. pennsylvanicus* males	
		More failed matings with heterospecifics	
Prevents sperm transfer	Mechanical	None	Harrison (1983)
Post-mating prezygotic			
Prevents gametes from meeting	Sperm transfer/storage	None	Present study
	Sperm viability	None	Present study
	Sperm utilization	?	
	Egg release	*Gryllus firmus* females mated with *G. pennsylvanicus* males lay few eggs	Harrison (1983), Maroja *et al.* (2009a)
	Sperm/egg attraction	?	
Prevents gametes from fusing	Sperm/egg interactions	?	
Prevents pronuclei fusing	Intracellular gametic incompatibility	?	
Post-zygotic			
Intrinsic	Hybrid inviability	A few eggs from *G. firmus* females mated with *G. pennsylvanicus* males appear to be fertilized, but these eggs fail to develop	Present study
		Hybrids from the reciprocal cross are viable	Harrison (1983)
	Hybrid sterility	None	Harrison (1983)
Extrinsic	Hybrid ecological inviability	?	
	Hybrid behavioural sterility	Hybrids are less likely to mate with *G. pennsylvanicus*	Maroja *et al.* (2009a)

References

Achmann, R., Heller, K.G. & Epplen, J.T. 1992. Last–male sperm precedence in the bushcricket *Poecilimon veluchianus* (Orthoptera, Tettigonioidea) demonstrated by DNA fingerprinting. Mol. Ecol. 1: 47–54.

Alexander, R. D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). Ann. Entomol. Soc. Am. 50: 584–602.

Andersson, M. 1994. Sexual Selection. Princeton University Press, Princeton, NJ, USA.

Andre´s, J.A., Maroja, L.S., Bogdanowicz, S.M., Swanson, W.J. & Harrison, R.G. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23: 1574–1584.

Andre´s, J.A., Maroja, L.S. & Harrison, R.G. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc. R. Soc. B 275: 1975–1983.

Arnqvist, G., Edvardsson, M., Friberg, U. & Nilsson, T. 2000. Sexual conflict promotes speciation in insects. Proc. Natl. Acad. Sci. USA 97: 10460–10464.

Backus, V.L. & Cade, W.H. 1986. Sperm competition in the field cricket *Gryllus integer* (Orthoptera, Gryllidae). Fla. Entomol. 69: 722–728.

Bella, J.L., Butlin, R.K., Ferris, C. & Hewitt, G.M. 1992. Asymmetrical homogamy and unequal sex ratio from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies. Heredity 68: 345–352.

Braswell, W.E., Andre´ s, J.A., Maroja, L.S., Harrison, R.G., Howard, D.J. & Swanson, W.J. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome 49: 1069–1080.

Bretman, A. & Tregenza, T. 2005. Measuring polyandry in wild populations: a case study using promiscuous crickets. Mol. Ecol. 14: 2169–2179.

Bretman, A., Newcombe, D. & Tregenza, T. 2009. Promiscuous females avoid inbreeding by controlling sperm storage. Mol. Ecol. 18: 3340–3345.

Broughton, R.E. & Harrison, R.G. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163: 1389–1401.

Bussie`re, L.F., Hunt, J., Jennions, M.D. & Brooks, R. 2006. Sexual conflict and cryptic female choice in the black field cricket, *Teleogryllus commodus*. Evolution 60: 792–800.

Civetta, A. & Singh, R.S. 1998. Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. 15: 901–909.

Clark, N.L., Aagaard, J.E. & Swanson, W.J. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131: 11–22.

Coyne, J.A. & Orr, H.A. 2004. Speciation. Sinauer Associates, Sunderland, MA, USA.

Crawley, M.J. 1993. GLIM for Ecologists. Blackwell Scientific Publishing, Oxford, UK.

Darwin, C. 1871. The Descent of Man, and Selection in Relation to Sex. J. Murray, London, UK.

Dieringer, D. & Schlo¨tterer, C. 2003. MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large micro- satellite data sets. Mol. Ecol. Notes 3: 167–169.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York, NY, USA.

Doherty, J.A. & Storz, M.M. 1992. Calling song and selective phonotaxis in the field crickets, *Gryllus firmus and G. pennsylvanicus* (Orthoptera, Gryllidae). J. Insect Behav. 5: 555–569.

Eady, P.E. 2001. Postcopulatory, prezygotic reproductive isolation. J. Zool. Lond. 253: 47–52.

Eberhard, W.G. 1996. Female Control: Sexual Selection by Cryptic Female Choice. Princeton University Press, Princeton, NJ, USA.

Evans, A.R. 1988. Mating systems and reproductive strategies in three Australian Gryllid crickets: *Bobilla victoriae* Otte, *Balamara gidya* Otte, and *Teleogryllus commodus* Walker (Orthoptera: Gryllidae: Nemobiinae; Trigonidiinae; Gryllinae). Ethology 78: 21–52.

Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403: 886–889.

Geyer, L.B. & Palumbi, S.R. 2005. Conspecific sperm precedence in two species of tropical sea urchins. Evolution 59: 97–105.

Gregory, G.E. 1965. The formation and fate of the spermatophore in the African migratory locust, *Locusta migratoria migratorioides* Reiche and Fairmaire. Trans. R. Ent. Soc. of Lond. 117: 33–66.

Gregory, P.G. & Howard, D.J. 1994. A postinsemination barrier to fertilization isolates two closely related ground crickets. Evolution 48: 705–710.

Gwynne, D.T. & Snedden, A.W. 1995. Paternity and female remating in *Requena verticalis* (Orthoptera, Tettigoniidae). Ecol. Entomol. 20: 191–194.

Hall, M.D., Bussie` re, L.F., Demont, M., Ward, P.I. & Brooks, R.C. 2010. Competitive PCR reveals the complexity of postcopulatory sexual selection in *Teleogryllus commodus*. Mol. Ecol. 19: 610–619.

Hamilton, M.B., Pincus, E.L., Di Fiore, A. & Fleischer, R.C. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques 27: 500–507.

Harper, F.M. & Hart, M.W. 2005. Gamete compatibility and sperm competition affect paternity and hybridization between sympatric *Asterias* sea stars. Biol. Bull. 209: 113–126.

Harrison, R.G. 1983. Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution 37: 245–251.

Harrison, R.G. 1985. Barriers to gene exchange between closely related cricket species. II. Life-cycle variation and temporal isolation. Evolution 39: 244–259.

Harrison, R.G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56: 337–349.

Harrison, R.G. & Arnold, J. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36: 535–552.

Harrison, R.G. & Bogdanowicz, S.M. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51: 493–505.

Harrison, R.G. & Rand, D.M. 1989. Mosaic hybrid zones and the nature of species boundaries. In: Speciation and its Consequences (D. Otte & J.A. Endler, eds), pp. 111–133. Sinauer, Sunder- land, MA, USA.

Helversen, D.V. & Helversen, O.V. 1991. Pre-mating sperm removal in the bushcricket *Metaplastes ornatus* Ramme 1931 (Orthoptera, Tettigonoidea, Phaneropteridae). Behav. Ecol. Sociobiol. 28: 391–396.

Hewitt, G.M., Mason, P. & Nichols, R.A. 1989. Sperm precedence and homogamy across a hybrid zone in the alpine grasshopper *Podisma pedestris*. Heredity 62: 343–353.

Hockham, L.R., Graves, J.A. & Ritchie, M.G. 2004. Sperm competition and the level of polyandry in a bushcricket with large nuptial gifts. Behav. Ecol. Sociobiol. 57: 149–154.

Holland, B. & Rice, W.R. 1998. Chase-away sexual selection: antagonistic seduction versus resistance. Evolution 52: 1–7.

Howard, D.J. 1999. Conspecific sperm and pollen precedence and speciation. Annu. Rev. Ecol. Syst. 30: 109–132.

Howard, D.J., Gregory, P.G., Chu, J. & Cain, M.L. 1998. Conspecific sperm precedence is an effective barrier to hybridization between closely related species. Evolution 52: 511–516.

Howard, D.J., Palumbi, S.R., Birge, L.M. & Manier, M.K. 2009. Sperm and speciation. In: Sperm Biology: An Evolutionary Perspective (T.R. Birkhead, D.J. Hosken & S. Pitnick, eds), pp. 267–403. Academic Press, Burlington, MA, USA.

Hunter-Jones, P. 1960. Fertilization of eggs of the desert locust by spermatozoa from successive copulations. Nature 185: 336– 336.

Kalinowski, S.T., Taper, M.L. & Marshall, T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16: 1099–1106.

Kawamura, N. 2001. Fertilization and the first cleavage mitosis in insects. Develop. Growth

Differ. 43:

Kelleher, E.S. & Markow, T.A. 2007. Reproductive tract interactions contribute to isolation in *Drosophila*. Fly 1: 33– 37.

Khalifa, A. 1950. Spermatophore production and egg-laying behavior in *Rhodnius prolixus* Stal. (Hemiptera, Reduviidae). Parasitology 40: 283–289.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. USA 78: 3721–3725.

Lo´ pez-Leo´n, M.D., Cabrero, J., Pardo, M.C., Viseras, E. & Camacho, J.P.M. 1993. Paternity displacement in the grass- hopper *Eyprepocnemis plorans*. Heredity 71: 539–545.

Lorch, P. D. & Servedio, M.R. 2005. Postmating-prezgyotic isolation is not an important source of selection for reinforcement within and between species in *Drosophila pseudoobscura* and *D. persimilis*. Evolution 59: 1039–1045.

Markow, T.A. 1997. Assortative fertilization in *Drosophila*. Proc. Natl. Acad. Sci. USA 94: 7756–7760.

Maroja, L.S., Clark, M.E. & Harrison, R.G. 2008. *Wolbachia* plays no role in the one-way reproductive incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*. Heredity 101: 435–444.

Maroja, L.S., Andre´s, J.A., Walters, J.R. & Harrison, R.G. 2009a. Multiple barriers to gene exchange in a field cricket hybrid zone. Biol. J. Linn. Soc. 97: 390–402.

Maroja, L.S., Andre´s, J.A. & Harrison, R.G. 2009b. Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. Evolution 63: 2999–3015.

Matute, D.R. & Coyne, J.A. 2010. Intrinsic reproductive isolation between two sister species of Drosophila. Evolution 64: 903–920.

Mayr, E. 1963. Animal Species and Evolution. Belknap Press, Cambridge, UK.

Morrow, E.H. & Gage, M.J.G. 2001. Sperm competition experiments between lines of crickets producing different sperm lengths. Proc. R. Soc. Lond. B 268: 2281–2286.

Mullen, S.P., Mendelson, T.C., Schal, C. & Shaw, K.L. 2007. Rapid evolution of cuticular hydrocarbons in a species radiation of acoustically diverse Hawaiian crickets (Gryllidae: Trigonidiinae: *Laupala*). Evolution 61: 223–231.

Nabours, R. K. 1927. Polyandry in the grouse locust, *Paratettix texanus* Hancock, with notes on inheritance of acquired characters and telegony. Am. Nat. 61: 531–538.

Ono, T., Siva-Jothy, M.T. & Kato, A. 1989. Removal and subsequent ingestion of rivals' semen during copulation in a tree cricket. Physiol. Entomol. 14: 195–202.

Panhuis, T.M., Butlin, R., Zuk, M. & Tregenza, T. 2001. Sexual selection and speciation. Trends Ecol. Evol. 16: 364– 371.

Parker, G.A. 1970. Sperm competition and its evolutionary consequences in insects. Biol. Rev. 45: 525–567.

Parker, G.A. & Partridge, L. 1998. Sexual conflict and speciation. Phil. Trans. R. Soc. Lond. B 353: 261–274.

Parker, G.A. & Smith, J.L. 1975. Sperm competition and the evolution of the precopulatory passive phase behaviour in *Locusta migratoria migratorioides*. J. Entomol. Ser. A. 49: 155– 171.

Pitnick, S., Markow, T. & Spicer, G.S. 1999. Evolution of multiple kinds of female sperm-storage organs in Drosophila. Evolution 53: 1804–1822.

Pitnick, S., Miller, G.T., Schneider, K. & Markow, T.A. 2003. Ejaculate-female coevolution in *Drosophila mojavensis*. Proc R. Soc. Lond. B 270: 1507–1512.

Rand, D. M. & Harrison, R. G. 1989. Ecological genetics of a mosaic hybrid zone – mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43: 432– 449.

Raymond, M. & Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86: 248–249.

Reinhardt, K. 2000. Variation in sperm precedence in *Chorthippus* grasshoppers (Caelifera: Gomphocerinae). Physiol. Entomol. 25: 324–329.

Reinhardt, K. & Meister, J. 2000. Low numbers of sperm retained in the spermatheca may explain the high values of sperm precedence in the migratory locust, *Locusta migratoria* (Latr.). J. Insect Behav 13: 839–849.

Rice, W.R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381: 232–234.

Rice, W.R. 1998. Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In: Endless Forms: Species and Speciation (D.J. Howard & S.H. Berlocher, eds), pp. 261–270. Oxford University Press, New York, NY, USA.

Ritchie, M.G. 2007. Sexual selection and speciation. Annu. Rev. Ecol. Evol. Syst. 38: 79–102.

Ross, C. L. & Harrison, R. G. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56: 2296–2312.

Ross, C. L. & Harrison, R. G. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115: 53–68.

Sakaluk, S.K. 1991. Post-copulatory mate guarding in decorated crickets. Anim. Behav. 41: 207–216.

Sakaluk, S.K. 1984. Male crickets feed females to ensure complete sperm transfer. Science 223: 609–610.

Sakaluk, S.K. 1986. Sperm competition and the evolution of nuptial feeding behavior in the cricket, *Gryllodes supplicans* (Walker). Evolution 40: 584–593.

Sakaluk, S.K. & Eggert, A.K. 1996. Female control of sperm transfer and intraspecific variation in sperm precedence: antecedents to the evolution of a courtship food gift. Evolution 50: 694–703.

Sarashina, I., Shinmyo, Y., Hirose, A., Miyawaki, K., Mito, T., Ohuchi, H. et al. 2003. Hypotonic buffer induces meiosis and formation of anucleate cytoplasmic islands in the egg of the two- spotted cricket *Gryllus bimaculatus*. Develop. Growth Differ. 45: 103–112.

Sarashina, I., Mito, T., Saito, M., Uneme, H., Miyawaki, K., Shinmyo, Y. et al. 2005. Location

of micropyles and early embryonic development of the two-spotted cricket *Gryllus bimaculatus* (Insecta, Orthoptera). Develop. Growth Differ. 47: 99–108.

Sato, M. & Tanaka-Sato, H. 2002. Fertilization, syngamy, and early embryonic development in the cricket *Gryllus bimaculatus* (De Geer). J. Morphol. 254: 266–271.

Simmons, L.W. 1986. Female choice in the field cricket *Gryllus bimaculatus* (De Geer). Anim. Behav. 34: 1463–1470.

Simmons, L.W. 1987. Sperm competition as a mechanism of female choice in the field cricket, *Gryllus bimaculatus*. Behav. Ecol. Sociobiol. 21: 197–202.

Simmons, L. W. 1995. Relative parental expenditure, potential reproductive rates, and the control of sexual selection in katydids. Am. Nat. 145: 797–808.

Simmons, L.W. 2001. Sperm Competition and its Evolutionary Consequences in the Insects. Princeton University Press, Princeton, NJ, USA.

Simmons, L.W. & Siva-Jothy, M.T. 1998. Sperm competition in insects: mechanisms and the potential for selection. In: Sperm Competition and Sexual Selection (T.R. Birkhead & A.P. Møller, eds), pp. 341–434. Academic press, San Diego, CA, USA.

Simmons, L.W., Wernham, J., Garc´ıa-Gonza´lez, F. & Kamien, D. 2003. Variation in paternity in the field cricket *Teleogryllus oceanicus*: no detectable influence of sperm numbers or sperm length. Behav. Ecol. 14: 539–545.

Snook, R.R., Chapman, T., Moore, P.J., Wedell, N. & Crudging- ton, H.S. 2009. Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. Evol. Ecol. 23: 71–91.

Solymar, B.D. & Cade, W.H. 1990. Heritable variation for female mating frequency in field crickets, *Gryllus integer*. Behav. Ecol. Sociobiol. 26: 73–76.

Sugawara, T. & Loher, W. 1986. Oviposition behaviour of the cricket *Teleogryllus commodus*: observation of external and internal events. J. Insect Physiol. 32: 179–188.

Swanson, W.J. & Vacquier, V.D. 2002. Reproductive protein evolution. Annu. Rev. Ecol. Syst. 33: 161–179.

Sweigart, A.L. 2010. The genetics of postmating, prezygotic reproductive isolation between *Drosophila virilis* and *D. americana*. Genetics 184: 401–410.

Thomas, M.L. & Simmons, L.W. 2009. Sexual selection on cuticular hydrocarbons in the Australian field cricket, *Telogryllus oceanicus*. BMC Evol. Biol. 9: 1471–2148.

Tregenza, T. & Wedell, N. 1997. Definitive evidence for cuticular pheromones in a cricket. Anim. Behav. 54: 979–984.

Vahed, K. 1998. Sperm precedence and the potential of the nuptial gift to function as paternal investment in the Tettigoniid *Steropleurus stali* Bolivar (Orthoptera: Tettigonidae: Epphippigerinae). J. Orthoptera Res. 7: 223–226.

Wedell, N. 1991. Sperm competition selects for nuptial feeding in a bushcricket. Evolution 45: 1975–1978.

Went, D.F. 1982. Egg activation and parthenogenetic reproduction in insects. Biol. Rev. 57: 319–344.

West-Eberhard, M.J. 1983. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58: 155–183.

Wilson, K. & Hardy, I.C.W. 2002. Statistical analysis of sex ratios: an introduction. In: Sex ratios: Concepts and Research Methods (I.C.W. Hardy, ed.) , pp. 48–92. Cambridge University Press, Cambridge, UK.

Wolfner, M.F. 2009. Battle and ballet: molecular inter- actions between the sexes in *Drosophila*. J. Hered. 100: 399–410.

Zhu, D.H. & Tanaka, S. 2002. Prolonged precopulatory mounting increases the length of copulation and sperm precedence in *Locusta migratoria* (Orthoptera : Acrididae). Ann. Entomol. Soc. Am. 95: 370–373.

Zuk, M. & Simmons, L.W. 1997. Reproductive strategies of the crickets (Orthoptera: Gryllidae). In: The Evolution of Mating Systems in Insects and Arachnids (J.C. Choe & B.J. Crespi, eds), pp. 89–109. Cambridge University Press, Cambridge, UK.

CHAPTER 2

INFLUENCE OF THE MALE EJACULATE ON POST-MATING PREZYGOTIC BARRIERS IN FIELD CRICKETS

Published as: Erica L. Larson, José A. Andrés, and Richard G. Harrison. 2012. Influence of the male ejaculate on post-mating prezygotic barriers in field crickets. PLoS ONE 7:e46202.

Abstract

Post-copulatory interactions between males and females involve highly coordinated, complex traits that are often rapidly evolving and divergent between species. Failure to produce and deposit eggs may be a common post-mating prezygotic barrier, yet little is known about what prevents the induction of egg-laying between species. The field crickets, *Gryllus firmus* and *G. pennsylvanicus* are isolated by a one-way reproductive incompatibility; *G. pennsylvanicus* males fail to fertilize *G. firmus* eggs or to induce normal egg-laying in *G. firmus* females. We use experimental crosses to elucidate the role of accessory gland-derived vs. testis-derived components of the *G. firmus* male ejaculate on egg-laying in conspecific and heterospecific crosses. Using surgical castrations to create 'spermless' males that transfer only seminal fluid proteins (SFPs) we test whether *G. firmus* male SFPs can induce egg-laying in conspecific crosses and rescue egg-laying in crosses between *G. pennsylvanicus* males and *G. firmus* females. We find *G. firmus* SFPs induce only a small short-term egg-laying response and that SFPs alone cannot explain the normal induction of egg-laying. *Gryllus firmus* SFPs also do not rescue the heterospecific cross. Testis-derived components, such as sperm or prostaglandins, most likely stimulate egg-laying or act as transporters for SFPs to targets in the female reproductive tract. These results highlight the utility of experimental approaches for investigating the phenotypes that act as barriers between species and suggest that future work on the molecular basis of the one-way incompatibility between *G. firmus* and *G. pennsylvanicus* should focus on divergent testis-derived compounds or proteins in addition to SFPs.

Introduction

Traits that mediate interactions between males and females are critical for reproduction and

yet often evolve rapidly and are highly divergent between species. Therefore, these traits may be particularly important in the early divergence of isolated populations and in speciation (West-Eberhard 1983; Andersson 1994; Civetta and Singh 1998; Panhuis et al. 2001; Ritchie 2007; Snook et al. 2009). Although we are often struck by the diversity of conspicuous behaviors involved in courtship and mate recognition, post-copulatory interactions between males and females are equally diverse and complex (Wolfner 1997, 2009). As a result, fertilization in a heterospecific cross can fail at a number of critical steps, resulting in post-mating prezygotic barriers between species. These barriers can range from traits that prevent sperm and eggs from meeting (e.g. sperm transfer, sperm storage, sperm utilization, egg-laying, sperm binding) to intracellular traits that prevent the sperm nucleus and egg nucleus from fusing (e.g. incomplete sperm entry, sperm folding) (Howard 1999; Howard et al. 2009; Snook et al. 2009; Larson et al. 2012).

Egg-laying in insects provides an example of the complexity of male-female interactions. Egg-laying is a multi-step process that involves egg production within the ovary (oogenesis), release of the egg from ovary into the oviducts (ovulation), progression of the egg down the oviducts, union of the sperm and egg within the genital chamber (fertilization) and the deposition of the egg into a particular substrate (oviposition). These steps are tightly linked to the proper transfer and storage of the male ejaculate. Oogenesis is increased when sperm is stored within the female storage organ, and ovulation interacts with sperm storage and sperm release from storage to facilitate successful fertilization (Figure 2.1, reviewed in Bloch Qazi et al. 2003). A reduction in the efficiency or a failure at any of these steps can lead to reproductive incompatibilities in insects.

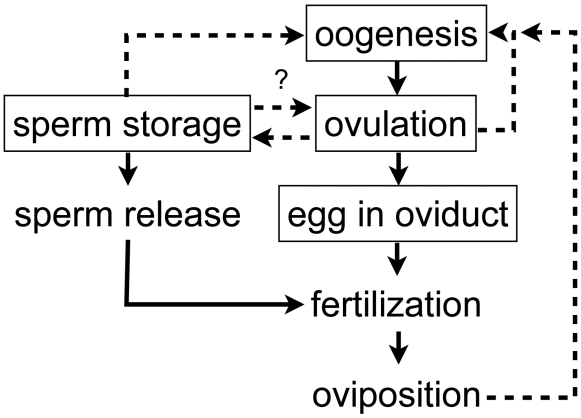


Figure 2.1 Egg-laying in female insects. Diagram showing the steps involved in egg-laying in female insects (solid lines) and the interactions that may stimulate increased egg-laying (dashed lines). In boxes are steps that may be mediated by seminal fluid proteins. The interaction between sperm storage and ovulation is hypothesized based on the observation that sperm transfer may increase ovulation and sperm depletion, thereby affecting the number of eggs laid. The figure is re-drawn from Bloch Qazi et al. (2003).

The failure of the male ejaculate to stimulate egg-laying between species has been observed in fruit flies (Fuyama 1983; Price et al. 2001), beetles (Wade et al. 1994), katydids (Shapiro 2000), lacewings (Albuquerque et al. 1996) and ground crickets (Gregory and Howard 1993). A similar phenomenon is sometimes observed between populations within a single species (Brown and Eady 2001; Messina et al. 2007). Failure to produce and deposit eggs may be a common post-mating prezygotic barrier, yet little is known about what prevents the induction of egg-laying between species. We do know which components of the male ejaculate (testis-derived vs. accessory gland- derived) induce intraspecific egg-laying in a variety of insect species (Leopold 1976), and the molecular interactions of the male ejaculate and female reproductive tract that induce egg-laying in *Drosophila melanogaster* are now well understood (Wolfner 1997; Bloch Qazi et al. 2003; Kubli 2003; Wolfner 2009). By characterizing the intraspecific mechanisms that result in egg-laying, we can begin to make inferences about how egg-laying breaks down between species. Here, we use experimental crosses both within and between species to test the influence of components of the male ejaculate on egg-laying and fertilization between two closely related species of field cricket.

Post-mating prezygotic barriers in field crickets

The field crickets, *Gryllus firmus* and *G. pennsylvanicus*, are recently diverged species (<0.5% mtDNA divergence (Willett et al. 1997)) that interact in a hybrid zone in the northeastern United States, extending from Massachusetts south into Virginia (Harrison and Arnold 1982). The cricket species have diverged both ecologically (Harrison 1985; Rand and Harrison 1989; Ross and Harrison 2002, 2006) and behaviorally (Harrison and Rand 1989; Doherty and Storz 1992; Maroja et al. 2009), but an important barrier between these species is a one-way

incompatibility between *G. firmus* females and *G. pennsylvanicus* males (Harrison 1983). Despite normal sperm transfer and storage, *G. firmus* females mated with *G. pennsylvanicus* males do not produce fertilized eggs (Maroja et al. 2008; Larson et al. 2012). Fertilization appears to break down in this cross somewhere between the release of sperm from storage and the sperm entering the egg (Larson et al. 2012). There is an equally striking reduction in egg-laying for these females. A *G. firmus* female mated with conspecifics will lay approximately 700 eggs over her lifetime, while a virgin female will produce less than 50 eggs and typically only late in life. *Gryllus firmus* females mated with *G. pennsylvanicus* males will lay about twice the number of eggs as virgin females, but significantly fewer eggs than a female mated to a conspecific (Maroja et al. 2009). In contrast, the reciprocal cross produces viable, fertile offspring in numbers indistinguishable from conspecific matings (Harrison 1983; Maroja et al. 2009).

Seminal fluid proteins (SFPs), which are synthesized and secreted from the male accessory gland and transferred to females during copulation, are known to play a role in many of the processes that may underlie a breakdown in egg-laying and fertilization, including ovulation, sperm storage, and sperm release (Figure 2.1). Accessory gland genes from *G. firmus* and *G. pennsylvanicus* have been characterized through transcriptome sequencing (Andrés et al. 2006; Braswell et al. 2006) and proteomics of the seminal fluid (Andrés et al. 2008). Many of the SFPs are found to be rapidly evolving under positive selection. However, there is currently no direct functional link between the divergence we observe in *Gryllus* SFPs and the post-mating prezygotic barriers that isolate these taxa. One step towards exploring this connection is to characterize the intraspecific and interspecific mechanism(s) that induce egg-laying and fertilization.

We attempt to elucidate the roles of accessory gland-derived vs. testis-derived components of the male ejaculate on these two barriers by asking whether *G. firmus* male SFPs induce egg-laying in *G. firmus* females and can "rescue" the cross between *G. firmus* females and *G. pennsylvanicus* males. We test the influence of *G. firmus* SFPs by mating females to surgically castrated ("spermless") conspecific males that transfer only SFPs. We find that SFPs induce only a modest short-term egg-laying response and that SFPs alone cannot explain the normal induction of egg-laying. We also try to rescue the incompatibility by mating *G. firmus* females to males of both species. Again, we find no evidence that *G. firmus* SFPs can rescue the one-way incompatibility.

Materials and Methods

Cricket Collections

We collected crickets in August of 2006, 2009 and 2011 from pure populations of *G. firmus* in Guilford, CT, USA (N 41°16'9"; W 72°39'59"); near Hammonasset Beach State Park, CT, USA (N 41°16'4"; W 72°34'14"); and Milford, CT, USA (N 41°11'48"; W 73°4'30") and *G. pennsylvanicus* in Ithaca, NY, USA (42°24'35"; -76°32'46"). Crickets were collected as late instar nymphs, separated by sex and maintained in large laboratory colonies with food (cat and rabbit food), water vials and egg flats for shelter, under a 12:12 h light/dark cycle at 28°C. Every two days we isolated crickets that had become adults and maintained them in same-sex groups of 6-8 crickets in plastic containers (30 x 16 x 9 cm). No specific permits were required for the described collections because the study organisms are not endangered or protected species and the collection locations are not privately own or protected.

Matings

For each experiment, virgin adult crickets between 6-10 days post-eclosion were randomly assigned to treatment (described below). We abbreviate treatments to indicate species (F = *G. firmus*, P = *G. pennsylvanicus*) and the order females were mated, with the first letter representing the female and subsequent letters representing the males with which she mated (*e.g.,* FFP represents a *G. firmus* female that mated with a *G. firmus* male followed by a *G. pennsylvanicus* male). Subscripts represent specific manipulations of male crickets described in the following section (*e.g.* FF_C represents a *G. firmus* female mated with a *G. firmus* male that was surgically castrated).

For each cross, cricket pairs were placed in petri dishes (9 cm) lined with moistened filter paper to provide traction. We considered matings complete when the male was observed to successfully transfer and properly attach the spermatophore to the female genital opening. To standardize the spermatophore attachment time and allow the spermatophore contents to be transferred completely, we left mated pairs undisturbed, allowing males to guard females and prevent early spermatophore removal. After males reinitiated courtship (approximately 45 min) we removed males from the mating chamber and females were either presented with a second male or were isolated in individual chambers depending on the experiment. Females that did not mate within 60 minutes of adding either a first or second male to the mating chamber were removed from the experiment.

Following matings, females were isolated in individual containers (30 x 16 x 9 cm) and provided with food, water, shelter and a petri dish (9 cm) filled with a mixture of moistened sand and soil as oviposition substrate. Food and water were replaced twice a week, oviposition substrate was periodically moistened, and mortality was scored every two days.

Surgical Castrations

To isolate the roles of testis-derived and accessory gland-derived components of the male ejaculate, we created spermless males using surgical castrations. Surgeries were performed on adult males 5-6 days post-eclosion, after the cuticle had hardened (surgeries performed too soon after eclosion result in high mortality rates). Males were divided into two categories: males that would be surgically castrated (F_C) and males that would undergo a sham castration (F_S) to serve as a control for effects of surgery on the male ejaculate production or content. Prior to surgery, males were anesthetized by chilling at 4^{o}C for at least 30 minutes. Using fine forceps, we made an incision across the dorsal side through the intersegmental membrane between the 2nd and 3rd abdominal segments and gently teased open the wound. For F_C males we completely removed each testis and severed the vas deferens (Figure 2.2), while for F_S males we probed the wound and body cavity to try to mimic testis removal. We then sealed the wound with Vetbond^{TM}Tissue Adhesive (3M, St. Paul, MN, USA), which polymerizes after contact with tissue and body fluids, binding the wound edges together. Following surgery, males were placed in a sterile petri dish with moistened cotton for water, and allowed two days to recover. Males in both categories had a high survival rate following surgery (F_C N = 83, 86.7% survived; F_S = 69, 89.8% survived). After the recovery period males resumed normal mate calling and courtship behaviors. Males were then transferred to individual containers and provided with food, shelter and water and the containers were cleaned every day.

Following the recovery period, males that were surgically castrated were placed in a petri dish with a single virgin *G. firmus* female, and allowed to mate repeatedly in order to deplete stores of mature sperm from the seminal vesicles. After two days of repeated matings, we checked males for remaining stored sperm by removing his spermatophore immediately after

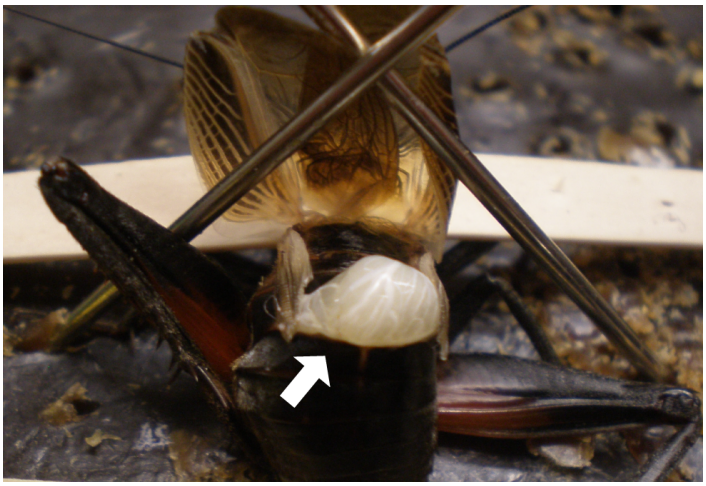


Figure 2.2 ***Gryllus firmus*** **male undergoing a surgical castration.** The incision is made between the dorsal 2nd and 3rd segments, and each testis (arrow) is gently removed.

attachment to the female and examining the spermatophore contents under a compound light microscope (400x). Spermatophores were gently removed so that all components of the spermatophore remained intact, placed on a microscope slide in a drop of phosphate buffered saline (PBS), and then gently squashed with a coverslip. When normal spermatophores were observed in this manner, a viscous liquid could be observed evacuating the spermatophore tube, followed immediately by long, thread-like sperm. If sperm were observed, males were allowed to mate repeatedly for another eight hours and were checked again the following day. If males were depleted of stored sperm, the viscous seminal fluid was still observed evacuating the spermatophore tube. If no sperm were observed in a male's spermatophore for three consecutive days, we considered the male spermless. Males were kept a minimum of 4 days during which they mated only once a day; this treatment allowed full recovery from repeated matings. Of the 72 males that survived surgical castrations, 57 were successfully cleared of sperm.

To validate that surgically castrated males still transferred SFPs we used two-dimensional electrophoresis (2D-E) to visualize the protein content of the spermless ejaculates. We collected two independent samples of both spermless spermatophores and normal spermatophores in liquid nitrogen, for a total of 4 samples (N = 25 males per sample). We homogenized each sample in 100 ul of ice-cold PBS and centrifuged (14,000 rpm for 1 min at 4^{o}C) to separate the ejaculate from most sperm and spermatophore debris. The 2D-E analysis, including sample preparation and quantification, first and second dimension separations using isoelectric focusing and Tris-SDS-PAGE electrophoresis, gel staining, image capture and analysis were carried out by the Cornell Core Laboratories Center for Proteomics and Mass Spectrometry.

Influence of the G. firmus ejaculate on egg-laying and longevity

To test the effects of seminal fluids on egg-laying in *G. firmus* females, we measured the total number of eggs produced by *G. firmus* females that 1) remained unmated (F = 57), 2) were mated to *G. firmus* males that were surgically castrated (FF_C = 56), 3) *G. firmus* males that underwent a sham castration (FF_S = 61), and 4) normal *G. firmus* males (FF = 60). By surgically castrating the *G. firmus* males, we could compare the effects on egg-laying of SFPs alone with the effects of the complete male ejaculate.

In other insects (including crickets) components of the male ejaculate typically elicit a short-term egg-laying response within 24 h of mating. To estimate both the initial egg-laying response and a female's lifetime fecundity, we collected oviposition substrates at both 48 h following mating and at the end of the female's lifespan. We collected substrates at 48 h to allow each female time to adjust following transfer to a new container and provide sufficient time for egg-laying; *G. firmus* females that are frequently disturbed are less inclined to oviposit (EL Larson personal observation). Eggs were separated from the oviposition substrate using a series of sieves and we counted the total number of eggs for each time point (within 48 h and after 48 h).

Influence of the G. firmus ejaculate on the one-way incompatibility

We performed two experiments to test whether the presence of a *G. firmus* male's ejaculate within the female reproductive tract could "rescue" the incompatibility (reduced egg-laying, no fertilization) between *G. firmus* females and *G. pennsylvanicus* males. For the first experiment *G. firmus* females were either mated to a normal *G. firmus* male immediately followed by a *G. pennsylvanicus* male (FFP = 9) or were mated first to a *G. pennsylvanicus* male followed by a

normal *G. firmus* male (FPF = 11). Females were provided with oviposition substrate immediately after mating, and then allowed to oviposit for three weeks. Eggs were incubated at 28°C for 21 days and then at 4°C for 102 days to break diapause conditions and ensure synchronous hatching (Harrison 1983). Eggs were then removed from chilled conditions and incubated at 28°C until hatching (approximately 17 days). We collected all hatchlings (1st instar nymphs) each morning until all eggs hatched and stored nymphs at -80°C for paternity analysis. We randomly selected 20 nymphs per cross for genotyping.

For paternity analysis, we used highly polymorphic microsatellite markers (PGI, Gr143, G3 and G28). Two of these loci were developed from *G. pennsylvanicus*, and have been previously described (Larson et al. 2012). The remaining two loci were developed from *G. firmus* using methods described in Hamilton et al. (1999) and Larson et al. (2012) with the addition of an enrichment by hybridization with biotinylated dimeric, trimeric and tetrameric nucleotide repeats. We quantified genetic variation for these new loci in one population of each species (*G. pennsylvanicus*: Ithaca, NY, USA; *G. firmus*: Guilford, CT, USA) (Table 2.1). Tests for linkage disequilibrium and deviation from Hardy-Weinberg equilibrium were performed using Genepop v. 4.1 (Raymond and Rousset 1995) and we adjusted significance thresholds using the false discovery rate procedure (Benjamini and Hochberg 1995). Cervus v 3.0 (Kalinowski et al. 2007) was used to test parentage exclusion probabilities, estimate null alleles and the polymorphic information content of the markers.

Parental genomic DNA extractions from single femurs were performed using the DNeasy Blood and Tissue Kit (QIAGEN Inc., Valencia, CA, USA); offspring genomic DNA was extracted from entire nymphs using the DNAdvance Genomic DNA Isolation Kit (Agencourt, Beverly, MA, USA). The forward primer of each primer pair was labeled with a 5' fluorescent

Table 2.1 Primer sequences and amplification conditions for *Gryllus* microsatellite loci used in paternity analysis.

Locus	Primer sequence (5′-3′)	T_a	Size	Sp	N	N_a	H_O	H_E	PIC	Null	GenBank
PGI	GAATGCATACATCAGTGTCATGAACA	56	220–	*Gf*	27	21	0.741	0.930	0.907	0.101	JN379460
(ATT_{15})	TGACTCAAAATAAGCATTATTTCAGC		334	*Gp*	14	15	0.929	0.939	0.898	0.011	
Gr143	CTGCCGCATTCACCAATCATTCAACTAT	58	150–	*Gf*	27	13	0.852	0.898	0.870	0.019	JN375328
(TG_{11})	CAACCAAGGGGCAAAATGAGTCAAACTT		204	*Gp*	14	9	0.857	0.820	0.763	0.038	
Gr3	GCGCGGCGACCGACTATTG	65–	153–	*Gf*	27	17	0.889	0.933	0.909	0.015	JX050157
(TG_{16})	CTCGCACCCTGTTAACAGTACTATCAAAAC	55	208	*Gp*	14	14	1.000	0.931	0.889	0.056	
Gr28	GCACCGCCCTAAACCCACGAC	65–	360–	*Gf*	27	6	0.667	0.648	0.762	0.046	JX050156
(TG_{11})	GGCACGGCAGCTTAAGGACATCAA	55	399	*Gp*	14	8	0.500	0.728	0.657	0.190	

T_a = annealing temperature (°C); Size = allele size range in base pairs; Sp = species; N = number of individuals scored; N_a = number of alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; PIC = polymorphic information content; Null = frequency of null alleles.
doi:10.1371/journal.pone.0046202.t001

tag (6-FAM, PET, NED, or VIC). We amplified these microsatellite loci using the Type-it Micosatellite PCR Kit (QIAGEN) following manufacturer's protocol with the addition of a touchdown protocol of 28 cycles of 95^{o}C for 30s, 59-53^{o}C for 90s (the annealing temperature decreased by 1^{o}C each cycle for the first 6 cycles and remained at 53^{o}C for the remaining 22 cycles) and 72^{o}C for 30s. Fluorescent PCR products were diluted 1:15 in water, mixed with formamide and GENESCAN LIZ-500 size standard (Applied Biosystems Inc. Foster City, CA, USA) and run on an ABI Automated 3730 DNA Analyzer at the Cornell University Life Sciences Core Laboratories Center (CLC). Alleles were called using GENEMAPPER (Applied Biosystems) and then verified by eye. The high level of polymorphism in our markers allowed us to assign paternity by eye.

For the second experiment *G. firmus* females were either mated to a surgically castrated male *G. firmus* male immediately followed by a *G. pennsylvanicus* male ($FF_CP = 3$), a *G. pennsylvanicus* male immediately followed by a surgically castrated *G. firmus* male ($FPF_C = 12$), or a normal *G. pennsylvanicus* male (FP = 10). After mating, females were provided with oviposition substrate and allowed to oviposit for 48 h. We then counted the total number of eggs laid by each female.

Statistics

To investigate the influence of *G. firmus* SFPs on *G. firmus* female egg-laying within 48 h of mating and after 48 h of mating, we constructed generalized linear mixed model (GLMM) contrasts to test the following hypotheses: Model 1: females mated with normal males (FF) will lay more eggs than females mated with males that that underwent sham castration (FF_S); Model 2: females mated with surgically castrated males (FF_C) will lay more eggs than unmated females

(F); and Model 3: females in treatments without sperm (F, FF_C) will lay fewer eggs than females in treatments with sperm (FF, FF_S). Differences between collecting locations were controlled by including a random effect of population identity. Using the R package 'lme4' (Bates et al. 2011), we modeled the proportion of females that laid eggs within 48 h with GLMMs fitted with a binomial error structure and a logit link function and the male treatment as the predictor. The egg-count data were highly over-dispersed; therefore, we modeled the total number of eggs laid by *G. firmus* females after 48 h using GLMMs with a Poisson distribution, individual-level random effects [42], and male treatment as the predictor. To compare longevity of females between mating treatments we estimated survival curves using the Kaplan-Meier method and compared differences between treatments using the log-rank test in the R package 'survival' (Therneau 2011).

To test the influence of the *G. firmus* ejaculate on the one-way incompatibility, we modeled the effect of male species on the proportion of offspring sired by the second male (P_2) for *G. firmus* females mated sequentially to both *G. firmus* and *G. pennsylvanicus* males. We constructed a GLM with binomial error structure and a logit link function using P_2 as the response variable and the male species as the predictor. For the second experiment involving *G. firmus* females mated with surgically castrated *G. firmus* males and *G. pennsylvanicus* males no statistics were required to interpret the results. Figures were constructed using the R packages 'plotrix' (Lemon 2006) and 'gplots' (Warnes 2010). All analyses were performed using the statistical package R version 2.12.0 (R Core Development Team, 2010).

Results

Seminal fluid protein content of spermless spermatophores

Representative examples of the 2D-E gels for the normal ejaculates (with sperm removed via centrifugation) and a spermless ejaculates are presented in Figure 2.3. Each spot represents a protein isoform. Overall, we estimate that there are about 630 protein spots present in all four samples. The patterns seen for normal and spermless spermatophore are very similar. Extra spots seen in the normal spermatophore extracts are presumably due to contamination from residual sperm.

Influence of the G. firmus ejaculate on egg-laying

When sperm are transferred to females (FF vs FF_S), there is no effect of sham castration on either the proportion of females that laid eggs within 48 h (Figure 2.4, GLMM: $z = -0.451$, $df = 6$, $p = 0.652$) or the number of eggs laid after 48 h (Figure 2.5, GLMM: $z = -1.12$, $df = 6$, $p = 0.263$). A greater proportion of females mated with surgically castrated males (FF_C) laid eggs in 48 h compared to females that remained unmated (F) (GLMM: $z = 2.442$, $df = 6$, $p = 0.015$); however, there was no significant difference in the number of eggs laid after 48 h (GLMM: $z = -1.38$, $df = 6$, $p = 0.167$). Comparisons between females mated with males that transferred sperm (FF, FF_S) and females that did not receive sperm (FF_C, F) revealed that in the former group there was both a greater proportion of females that laid eggs within 48 h (GLMM: $z = 7.155$, $df = 6$, $p = <0.001$) and females laid more eggs after 48 h (GLMM: $z = -10.15$, $df = 6$, $p = <0.001$).

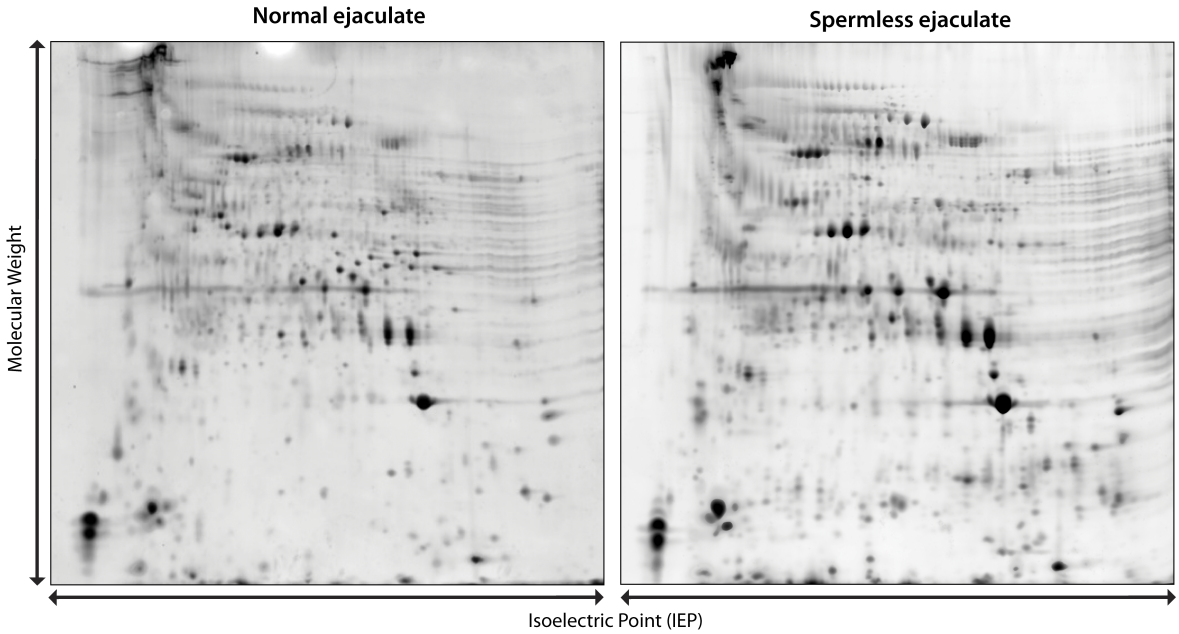


Figure 2.3 Two-dimensional gel electrophoresis of male ejaculates. Protein gels of a normal male ejaculate (left) and a spermless male ejaculate (right). Spermatophore samples were ground and centrifuged to remove spermatophore debris and sperm. Proteins were separated based on their isoelectric point in the first dimension and molecular weight in the second.

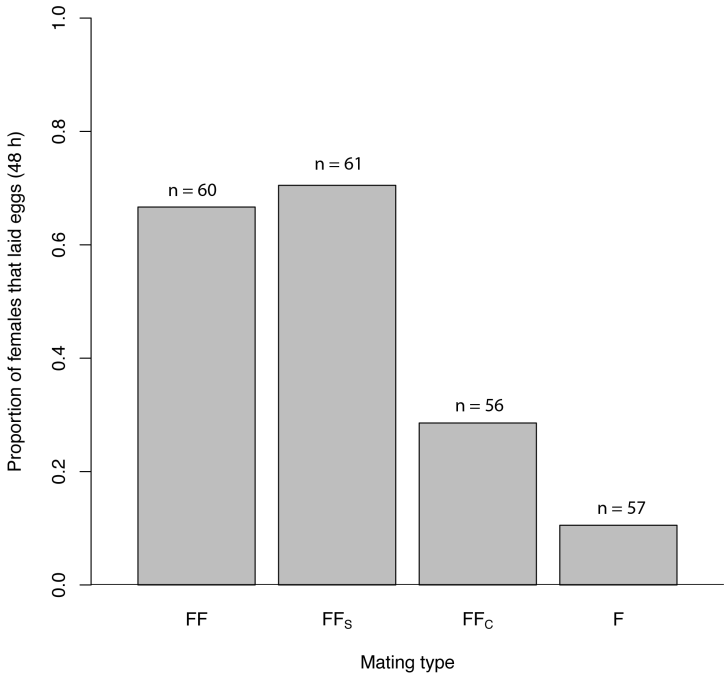


Figure 2.4 Proportion of *G. firmus* females that laid eggs within 48 h of mating. Females were mated with (1) a normal *G. firmus* male (FF), (2) a *G. firmus* male that underwent sham testes removal surgery (FF_S), (3) a *G. firmus* male surgically castrated (FF_C) or (4) remained unmated (F).

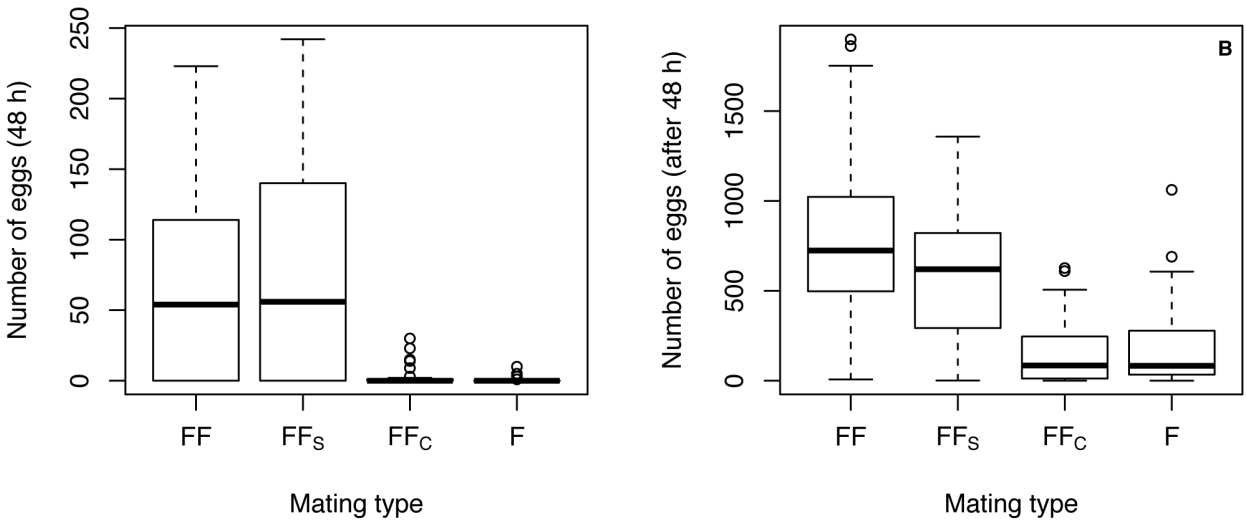


Figure 2.5 ***Gryllus firmus*** **female egg production.** Box plot of egg production **A)** 48 h following mating and **B)** total lifetime for *G. firmus* females that (1) mated with a normal *G. firmus* male (FF), (2) mated with a *G. firmus* male that underwent sham testes removal surgery (FF_S), (3) mated with a *G. firmus* male surgically castrated (FF_C) or (4) remained unmated (F).

Influence of the G. firmus ejaculate on female longevity

Female life span ranged from 7-84 days following mating (FF: 7- 71; FF_S: 17-84; FF_C: 17-69; F: 17-73) with an average lifespan of 44 days (FF: 45.8; FF_S: 41.8; FF_C: 42.7; 47.6). There was no difference in lifespan among the four mating treatments (Figure 2.6, log-rank: $\chi^2 = 0.2$, df = 3, p = 0.972).

Influence of the G. firmus ejaculate on one-way incompatibility

None of the microsatellite loci used in this study deviated from Hardy-Weinberg equilibrium following false discovery rate correction and there was no evidence of linkage disequilibrium between any pair of loci (Table 2.1). Despite the presence of null alleles in two of these loci (PGI and G28), the combined nonexclusion probability of the second parent across all four loci was 0.004 and 0.006 for *G. firmus* and *G. pennsylvanicus,* respectively, indicating that these markers are appropriate for assigning paternity. Of the 474 nymphs selected for genotyping, 98.9% were genotyped and assigned paternity successfully. The remaining 1.1% had poor microsatellite amplification, most likely due to low quantities of DNA as a result of little starting material. We did not observe any non-parental alleles in the offspring. *Gryllus firmus* females mated sequentially to *G. firmus* and *G. pennsylvanicus* males (FFP, FPF) produced offspring that were sired only by *G. firmus* males regardless of mating order ($t = -8.05 \times 10^{15}$, df = 1, p = <0.001). *Gryllus firmus* females mated sequentially to surgically castrated *G. firmus* males and *G. pennsylvanicus* males (FF_CP, FPF_C) and *G. firmus* females mated only to *G. pennsylvanicus* males (FP) laid no eggs within a 48 h period.

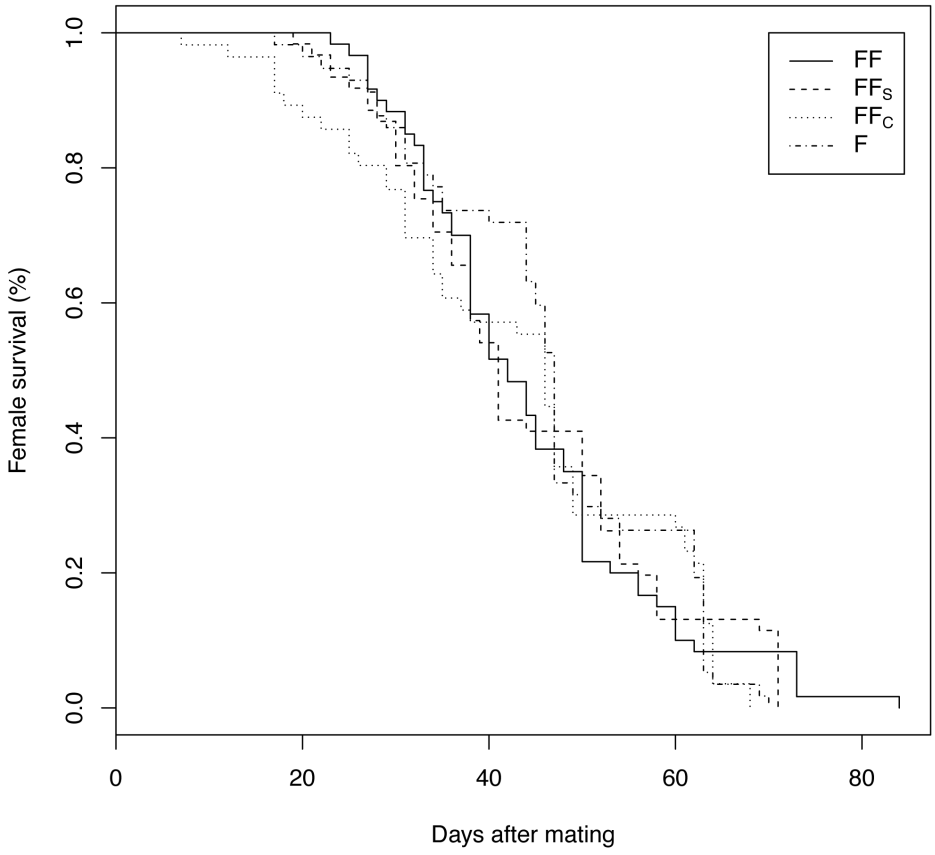


Figure 2.6 ***Gryllus firmus*** **female longevity**. Survivorship curves of *G. firmus* females (1) mated with a normal *G. firmus* male (FF), (2) mated with a *G. firmus* male that underwent sham testes removal surgery (FF_S), (3) mated with a *G. firmus* male surgically castrated (FF_C) or (4) remained unmated (F)

Discussion

Gryllus firmus *seminal fluid proteins induce only a marginal egg-laying response*

Following mating, female insects undergo numerous physiological and behavioral changes. Pre-mating courtship or the mechanical stimulus of copulation can cause these changes, but the majority of changes are induced by components of the male ejaculate. In insects, sperm clearly play a critical role in fertilization, but many aspects of sperm function (viability, storage, activation, competition) and female response to mating (mating refractoriness, oogenesis, sperm utilization, ovulation, oviposition) are mediated by SFPs secreted from the male accessory glands or ejaculatory duct (Engelmann 1970; Leopold 1976; Wolfner 1997; Gillott 2003; Wolfner 2009). Seminal fluid proteins were first linked to these post-mating changes via whole tissue transplantation experiments in *Drosophila melanogaster*, where portions of the male accessory glands or testes were transplanted into the abdomens of virgin females (Garcia-Bellido 1964; Merle 1968). Subsequently, methods to isolate the roles of SFPs in fertilization have ranged from simple injections of SFP extracts into virgin females to the creation of "spermless" or "accessory glandless" males and the targeted knockdown of specific SFPs using RNAi (Wolfner 1997; Gillott 2003; Wolfner 2009).

Much of this effort has focused on how components of the male ejaculate elevate egg-laying in mated females. In *Drosophila*, where egg-laying is best understood, the male ejaculate alters a female's reproductive physiology over different timescales. Initially, there is a short-term increase in the number of eggs laid within the first 24 h of mating (Kalb et al. 1993; Herndon and Wolfner 1995). This short-term response is induced by the presence of at least three SFPs in the female reproductive tract, sex peptide (SP, Acp70A), the prohormone ovulin (Acp26Aa) and CG33943 (Herndon and Wolfner 1995; Heifetz et al. 2000; Chapman et al. 2003;

Liu and Kubli 2003; Ram and Wolfner 2007). However, this response is transient, and the presence of both sperm and SFPs in the sperm storage organ is required to maintain elevated levels of egg-laying (Manning 1962, 1967). This so-called sperm effect or long-term post-mating response was thought to be induced by sperm-binding receptors or stretch receptors within the female sperm storage organ (Manning 1962). It is now clear that the sperm effect, at least in *Drosophila*, is actually an SFP effect, mediated in part by the SFP sex peptide (Chapman et al. 2003; Liu and Kubli 2003). Sex peptide binds to the tails of sperm and is slowly released from sperm within the female storage organ (Kubli 1992; Kubli 2003; Peng et al. 2005). It appears that sperm may act as both carriers and reservoirs for SFPs, enabling sperm to reach target cells within the female reproductive tract and maintaining their effects on female reproduction over an extended period. At least four other SFPs have been identified that act in concert with sex peptide to sustain the long-term post-mating response, and at least one of these proteins also binds to sperm (Ram and Wolfner 2007; Ram and Wolfner 2009).

Drosophila has been a model for understanding post-mating male and female interactions, and from this work it has become clear that seminal fluid proteins stimulate egg-laying in mated females but that to do so they must interact with sperm. However, this picture of reproduction appears to vary greatly across taxa. In other Dipterans there is evidence that SFPs alone can induce egg-laying in mated females (Leahy and Craig 1965; Thailayil et al. 2011), whereas in Lepidoptera egg-laying is often triggered by the presence of eupyrene sperm in the spermatheca (Thibout 1979; Karube and Kobayashi 1999; Xu and Wang 2011), but there is at least one case of SFPs inducing partial egg-laying (Jin and Gong 2001). There are few examples from the Coleoptera, but in at least two species both components of the testis/seminal vesicle and the accessory gland induce egg-laying, although the accessory gland extracts had a minimal

influence (Yamane and Miyatake 2010).

In Orthoptera, the picture is even less clear. Egg-laying is stimulated by SFPs in some grasshoppers (Pickford et al. 1969; Leahy 1973; Friedel and Gillott 1976; Lange and Loughton 1985; Yi and Gillott 1999) and in ground crickets (Marshall et al. 2009), while in at least one grasshopper the combination of mechanical stimulus and testis derived components can induce egg-laying (Quo 1959). In the field crickets, *Acheta domesticus* and *Teleogryllus commodus* egg-laying is initially induced by prostaglandins, autocrine hormones transferred to females as part of the seminal fluid (Destephano and Brady 1977; Loher 1979; Loher et al. 1981; Lange 1984; Stanley-Samuelson and Peloquin 1986), but the presence of sperm in the spermatheca is required to maintain long term egg-laying (similar to *Drosophila*) (Murtaugh and Denlinger 1985; Murtaugh and Denlinger 1987). Prostaglandins or prostaglandin precursors have been found to be synthesized in both the testes (*T. commodus* and *A. domesticus*) and the accessory glands (*A. domesticus* and *Locusta migratoria*) of Orthoptera, although the prostaglandins found in *L. migratoria* do not appear to be involved in egg-laying (Stanley and Kim 2011). The only study to attempt to induce egg-laying in field crickets using whole ejaculatory-fluid extracts failed to see a response (Green and Tregenza 2009). However, in that study ejaculate extracts were injected into the abdominal cavity and may have failed to elicit an egg-laying response because SFPs did not reach target receptors within the female reproductive tract.

Our use of castrated males to transfer SFPs to virgin females is a more effective way of delivering SFPs directly into the female reproductive tract while controlling for any effects of mating. Still, we found that SFPs without the presence of sperm or other testis-derived compounds induced only a modest short-term egg-laying response in the field cricket *G. firmus*. This response is small compared to egg-laying in normally mated females, a result in stark

contrast to the induction of egg-laying seen in other taxa. In the long-term, there was no difference in the fecundity of virgin females and those that received SFPs. This suggests that SFPs indeed play some role in eliciting egg-laying behavior over the short term, but that testis derived factors are required for both the short and long term post-mating egg-laying response in *G. firmus*.

It might be argued that the "effectors" of egg-laying did not reach their targets in the female reproductive tract. However, we know that spermatophores of surgically castrated males transfer seminal fluid. Seminal fluid can be directly observed evacuating the spermatophore tube of spermless spermatophores (see Methods). Furthermore, analysis of this fluid using 2-D gel electrophoresis clearly reveals the same pattern of protein spots that are seen in extracts from normal spermatophores (Figure 2.3).

There is also no reason to believe that the multiple matings required to create the spermless males adversely affect SFP volume or content. Field crickets are highly promiscuous and males mate repeatedly both in the wild and the laboratory (Alexander and Otte 1967; Sakaluk 1987; Simmons 1988; Tregenza and Wedell 1998; Wagner et al. 2001; Sakaluk et al. 2002; Wagner and Harper 2003; Bretman and Tregenza 2005; Rodriguez-Munoz et al. 2010). Males will only re-mate when a fully formed spermatophore is present in their spermatophore pouch (Loher and Dambach 1989; Zuk and Simmons 1997). As a result, both the timing and the frequency of matings in field crickets are dependent on spermatophore production, which in *Gryllus firmus* males is approximately every 45 minutes throughout the day (Alexander and Otte 1967; Maroja et al. 2009). Both the spermatophore and the seminal fluid are composed of proteins secreted by the male accessory gland (Chapman 1998; Heller et al. 1998) and it is unlikely that a male would have sufficient accessory gland function to produce a spermatophore, but not the seminal fluid

proteins. There is also evidence that male insect ejaculate content is consistent across repeated matings (Schaus and Sakaluk 2002) and throughout their lifetime (Chapman and Wolfner 1988; Monsma et al. 1990). In our protocol, castrated males were mated repeatedly only during the first two days following surgery recovery. Subsequently, they only mated once a day for a minimum of four days, well below the expected number of matings for a male field cricket. Thus, our treatment should not compromise SFP production.

Given our observation of SFPs in the seminal fluid of spermless spermatophores and our delivery method of SFPs directly into the female reproductive tract (as opposed to abdominal injections), our failure to find any large or long term egg-laying response induced by SFPs suggests that SFPs alone are not sufficient to stimulate egg-laying in *G. firmus*. This is consistent with a similar failure of SFPs to induce egg-laying in *G. bimaculatus* (Green and Tregenza 2009). It is possible, even very likely, that we see a failure of SFPs to induce egg-laying in field crickets because key SFPs bind to the sperm for transport into the female reproductive tract as has been demonstrated in *Drosophila* (Kubli 1992; Neubaum and Wolfner 1999; Chapman et al. 2003; Kubli 2003; Liu and Kubli 2003; Ram and Wolfner 2009).

The question of what components of the male ejaculate stimulate egg-laying is an important one, not simply for a better understanding of insect reproduction, but because these components, if diverged, may constitute a barrier to gene exchange between closely related taxa. To our knowledge, only one study, in *Drosophila pulchrella* and *D. suzukii,* has attempted to differentiate between the components of the male ejaculate that induce egg-laying in heterospecific crosses. In that case, a one-way incompatibility between *D. pulchrella* females and *D. suzukii* males is a result of both low sperm storage and severely reduced egg-laying. When *D. pulchrella* females are implanted with accessory gland tissue from conspecifics they

have an ovulation rate that is 75% of a normally mated female, whereas females implanted with heterospecific accessory glands have an ovulation rate of only 54% (Fuyama 1983; Ohashi et al. 1991).

Gryllus firmus seminal fluids do not affect female lifespan

Mating is often costly to females and results in decreased lifespan due to SFPs that are toxic to females. For example in *Drosophila,* females that receive SFPs during mating have a reduced lifespan (Chapman et al. 1995), but these SFPs serve to increase male mating success (Clark et al. 1995). In species, such as crickets, that are promiscuous and mate more often than is required for fertilization of their eggs, the male ejaculate may actually increase female lifespan (Wagner et al. 2001), or have no effect (Bateman et al. 2006). In one case, female lifespan in the cricket *G. bimaculatus* was reduced as a result of the injection of SFPs into the female abdomen, but it is difficult to determine whether this is a normal effect of SFPs or a result of SFPs present in the body cavity where they may be toxic (Green and Tregenza 2009). We found no effect of SFPs on female lifespan. This is consistent with similar studies in *Gryllus firmus* and *G. pennsylvanicus* that found no difference in the lifespan of singly mated, doubly mated and virgin females (Maroja et al. 2009), despite the fact that virgin females and females mated with surgically castrated males are often 'bursting' with eggs (EL Larson personal observation). It is possible that there are lifespan benefits or costs to mating, but that a greater number of matings is required to see an effect in *G. firmus*.

Gryllus firmus *seminal fluid proteins alone do not rescue the one-way incompatibility*

In *Drosophila*, sperm function is dependent on the presence of SFPs, and males that

transfer only sperm (*prd* males) are completely sterile. However, females mated to *prd* males and subsequently mated to males transferring only SFPs (*tud* males) can occasionally lay fertile eggs (Xue and Noll 2000). The number of rescued crosses is very low (less than 1%). Nonetheless, this suggests that SFPs from one male can facilitate fertilization by the sperm of a second male. Alternatively, SFPs could act as a specific stimulus, only affecting sperm from the same male, or complementation may only be possible within species and SFPs may not interact with heterospecific sperm.

Our results suggest that in *G. firmus*, the latter is the case. The *G. firmus* male ejaculate was unable to facilitate fertilization for heterospecific sperm and *G. firmus* males sired all offspring in females mated sequentially with both species. Similar results have been observed in double matings of the lacewing species *Chrysopa quadripunctata* and *C. slossonae*, but in these taxa, there are fewer heterospecific sperm stored (Albuquerque et al. 1996). *Gryllus firmus* male SFPs also failed to induce egg-laying when *G. pennsylvanicus* sperm were present in the spermatheca. Therefore, failure of *G. firmus* SFPs alone to stimulate normal egg-laying in *G. firmus* females is not simply a result of mechanical stimulus (e.g., stretch receptors in the spermatheca). SFPs may need to act in concert with sperm to induce egg-laying in *G. firmus* females. Unfortunately, we are not able to test whether testis-derived compounds alone can induce oviposition. Surgical removal of accessory glands would prevent the formation of the spermatophore necessary to transfer sperm to a female.

Conclusions

Our results highlight the utility of experimental approaches for investigating the phenotypes that act as barriers between species and provide new directions for investigating the

molecular changes that lead to these barriers. The nature of the one-way incompatibility between *G. firmus* females and *G. pennsylvanicus* males suggests a role for SFPs, and both egg-laying and fertilization are traits that are often mediated by SFPs in other taxa. In addition, many SFPs are highly divergent between *G. firmus* and *G. pennsylvanicus* and appear to be evolving as a result of positive selection (Andrés et al. 2006; Andrés et al. 2008). Although the results of this study do not exclude a role for SFPs in these barriers, they do suggest that SFPs are not solely responsible for successful egg-laying. In particular, testis derived components, such as sperm or prostaglandins, either stimulate egg-laying or act as transporters for SFPs to targets in the female reproductive tract. Future work on the molecular basis of the one-way incompatibility between *G. firmus* and *G. pennsylvanicus* should focus on divergent testis-derived compounds or proteins.

A great deal of research on post-mating prezygotic barriers in internal fertilizers has concentrated on the role of SFPs (Swanson and Vacquier 2002; Clark et al. 2006; Turner and Hoekstra 2008), but our results suggest that focusing on SFPs alone is too narrow. Although there are now numerous examples of rapid divergence in SFPs between closely related species across diverse taxonomic groups (Turner and Hoekstra 2008) and there is some evidence of post-mating prezygotic barriers between several of these species (Andrés et al. 2008; Marshall et al. 2011), there are few studies that provide a functional link between the rapid evolution of SFPs and post-mating prezygotic barriers. While documenting patterns of divergence between species is an important step, functional studies through experimental crosses are needed to determine whether divergent genes play a role in reproductive barriers between species.

Acknowledgements

We thank Nicholas Saleh and Eliana Bondra for assistance with data collection, Joseph Simonis for assistance with statistical analyses and members of the Harrison laboratory for thoughtful comments on the manuscript.

References

Albuquerque, G., C. Tauber, and M. Tauber. 1996. Postmating reproductive isolation between *Chrysopa quadripunctata* and *Chrysopa slossonae*: Mechanisms and geographic variation. Evolution 50:1598-1606.

Alexander, R., and D. Otte. 1967. The evolution of genitalia and mating behavior in crickets (Gryllidae) and other Orthoptera. Misc Publ Mus Zool Univ Mich 133:1-62.

Andersson, M. 1994. Sexual selection. Princeton University Press, Princeton.

Andrés, J. A., L. S. Maroja, S. M. Bogdanowicz, W. J. Swanson, and R. G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23:1574-1584.

Andrés, J. A., L. S. Maroja, and R. G. Harrison. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 275:1975-1983.

Bateman, P. W., J. W. H. Ferguson, and C. A. Yetman. 2006. Courtship and copulation, but not ejaculates, reduce the longevity of female field crickets (*Gryllus bimaculatus*). J. Zool. 268:341-346.

Bates, D., M. Maechler, and B. Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. R Foundation for Statistical Computing, Vienna, Austria.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Ser. B. (Stat. Method.) 57:289-300.

Bloch Qazi, M., Y. Heifetz, and M. F. Wolfner. 2003. The developments between gametogenesis and fertilization: ovulation and female sperm storage in *Drosophila melanogaster*. Dev. Biol. 256:195-211.

Braswell, W. E., J. A. Andres, L. S. Maroja, R. G. Harrison, D. J. Howard, and W. J. Swanson. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome 49:1069-1080.

Bretman, A., and T. Tregenza. 2005. Measuring polyandry in wild populations: a case study using promiscuous crickets. Mol. Ecol. 14:2169-2179.

Brown, D. V., and P. E. Eady. 2001. Functional incompatibility between the fertilization systems of two allopatric populations of *Callosobruchus maculatus* (Coleoptera : Bruchidae). Evolution 55:2257-2262.

Chapman, K. B., and M. F. Wolfner. 1988. Determination of male-specific gene expression in *Drosophila* accessory glands. Dev. Biol. 126:195-202.

Chapman, R. F. 1998. The Insects: Structure and Function. Cambridge University Press, New York.

Chapman, T., J. Bangham, G. Vinti, B. Seifried, O. Lung, M. F. Wolfner, H. K. Smith, and L. Partridge. 2003. The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. Proc. Natl. Acad. Sci. U. S. A. 100:9923-9928.

Chapman, T., L. F. Liddle, J. M. Kalb, M. F. Wolfner, and L. Partridge. 1995. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature 373:241-244.

Civetta, A., and R. S. Singh. 1998. Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. 15:901-909.

Clark, A. G., M. Aguade, T. Prout, L. G. Harshman, and C. H. Langley. 1995. Variation in sperm displacement and its association with accessory-gland protein loci in *Drosophila-melanogaster*. Genetics 139:189-201.

Clark, N. L., J. E. Aagaard, and W. J. Swanson. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131:11-22.

Destephano, D. B., and U. E. Brady. 1977. Prostaglandin and prostaglandin synthetase in the cricket, *Acheta domesticus*. J. Insect Physiol. 23:905-911.

Doherty, J. A., and M. M. Storz. 1992. Calling song and selective phonotaxis in the field crickets, *Gryllus firmus* and *G. pennsylvanicus* (Orthoptera, Gryllidae). J. Insect Behav. 5:555-569.

Engelmann, F. 1970. The physiology of insect reproduction. Pergamon Press.

Friedel, T., and C. Gillott. 1976. Male accessory gland substance of *Melanoplus sanguinipes*: ovipostion stimulant under control of corpus allatum. J. Insect Physiol. 22:489-&.

Fuyama, Y. 1983. Species-specificity of paragonial substances as an isolating mechanism in *Drosophila*. Experientia 39:190-192.

Garcia-Bellido, A. 1964. Das sekret der paragonien als stimulus der fekunditat bei weibchen von *Drosophila melanogaster*. Zeitschrift Fur Naturforschung Part B-Chemie Biochemie Biophysik Biologie Und Verwandten Gebiete B 19:491- 495.

Gillott, C. 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. Annu. Rev. Entomol. 48:163-184.

Green, K., and T. Tregenza. 2009. The influence of male ejaculates on female mate search behaviour, oviposition and longevity in crickets. Anim. Behav. 77:887-892.

Gregory, P. G., and D. J. Howard. 1993. Laboratory hybridization studies of *Allonemobius fasciatus* and *A. socius* (Orthoptera, Gryllidae). Ann. Entomol. Soc. Am. 86:694-701.

Hamilton, M. B., E. L. Pincus, A. Di Fiore, and R. C. Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques 27:500-507.

Harrison, R., and S. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36:535-552.

Harrison, R. G. 1983. Barriers to gene exchange between closely related cricket species . I. laboratory hybridization studies. Evolution 37:245-251.

Harrison, R. G. 1985. Barriers to gene exchange between closely related cricket species . II. life-cycle variation and temporal isolation. Evolution 39:244-259.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 111 - 133 *in* D. Otte, and J. A. Endler, eds. Speciation and its Consequences. Sinauer, Sunderland, MA.

Heifetz, Y., O. Lung, E. A. Frongillo Jr, and M. F. Wolfner. 2000. The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Curr. Biol. 10:99-102.

Heller, K. G., S. Faltin, P. Fleischmann, and O. von Helversen. 1998. The chemical composition of the spermatophore in some species of phaneropterid bushcrickets (Orthoptera : Tettigonioidea). J. Insect Physiol. 44:1001-1008.

Herndon, L. A., and M. F. Wolfner. 1995. A *Drosophila* seminal fluid protein, Acp26Aa, stimulate egg-laying in females for one day after mating. Proc. Natl. Acad. Sci. U. S. A. 92:10114-10118.

Howard, D. J. 1999. Conspecific sperm and pollen precedence and speciation. Annu. Rev. Ecol. Syst. 30:109-132.

Howard, D. J., S. R. Palumbi, L. M. Birge, and M. K. Manier. 2009. Sperm and speciation. Pp. 267-403 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm Biology: An Evolutionary Perspective. Academic Press, Burlington, MA, USA.

Jin, Z.-Y., and H. Gong. 2001. Male accessory gland derived factors can stimulate oogenesis and enhance oviposition in *Helicoverpa armigera* (Lepidoptera: Noctuidae). Arch. Insect Biochem. Physiol. 46:175-185.

Kalb, J. M., A. J. Dibenedetto, and M. F. Wolfner. 1993. Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. Proc. Natl. Acad. Sci. U. S. A. 90:8093-8097.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099-1106.

Karube, F., and M. Kobayashi. 1999. Presence of eupyrene spermatozoa in vestibulum accelerates oviposition in the silkworm moth, *Bombyx mori*. J. Insect Physiol. 45:947-957.

Kubli, E. 1992. My favorite molecule. The sex-peptide. Bioessays 14:779-784.

Kubli, E. 2003. Sex-peptides: seminal peptides of the *Drosophila* male. Cell. Mol. Life Sci. 60:1689-1704.

Lange, A. B. 1984. The transfer of prostaglandin-synthesizing activity during mating in *Locusta migratoria*. Insect Biochem. 14:551-556.

Lange, A. B., and B. G. Loughton. 1985. An oviposition-stimulating factor in the male accessory reproductive gland of the locust, *Locusta migratoria*. Gen. Comp. Endocrinol. 57:208-215.

Larson, E. L., G. L. Hume, J. A. Andrés, and R. G. Harrison. 2012. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. J. Evol. Biol. 25:174-186.

Leahy, M. G. 1973. Oviposition of virgin *Schistocerca gregaria* (Forskal) (Orthoptera, Acrididae) after implant of male accessory gland complex. J Entomol Ser A Physiol Behav 48:69-78.

Leahy, M. G., and G. B. Craig. 1965. Accessory gland substance as a stimulant for oviposition in *Aedes aegypti* and *Aedes albopictus*. Mosq News 25:448-452.

Lemon, J. 2006. Plotrix: a package in the red light district of R. R-News 6:8-12.

Leopold, R. A. 1976. Role of male accessory glands in insect reproduction. Annu. Rev. Entomol. 21:199-221.

Liu, H. F., and E. Kubli. 2003. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 100:9929-9933.

Loher, W. 1979. The influence of prostaglandin E2 on oviposition in *Teleogryllus commodus*. Entomol. Exp. Appl. 25:107-109.

Loher, W., and M. Dambach. 1989. Reproductive Behavior. Pp. 43-82 *in* F. Huber, T. E. Moore, and W. Loher, eds. Cricket Behavior and Neurobiology. Cornell University Press, Ithaca, NY.

Loher, W., I. Ganjian, I. Kubo, D. Stanleysamuelson, and S. S. Tobe. 1981. Prostaglandins - their role in egg-laying of the cricket *Teleogryllus commodus* Proc. Natl. Acad. Sci. U. S. A. 78:7835-7838.

Manning, A. 1962. Sperm factor affecting receptivity of *Drosophila melanogaster* females. Nature 194:252-253.

Manning, A. 1967. Control of sexual receptivity in female *Drosophila*. Anim. Behav. 15:239-250.

Maroja, L. S., J. A. Andrés, J. R. Walters, and R. G. Harrison. 2009. Multiple barriers to gene exchange in a field cricket hybrid zone. Biol. J. Linn. Soc. 97:390-402.

Maroja, L. S., M. E. Clark, and R. G. Harrison. 2008. *Wolbachia* plays no role in the one-way reproductive incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*. Heredity 101:435-444.

Marshall, J. L., D. L. Huestis, C. Garcia, Y. Hiromasa, S. Wheeler, S. Noh, J. M. Tomich, and D. J. Howard. 2011. Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. Mol. Biol. Evol. 28:423-435.

Marshall, J. L., D. L. Huestis, Y. Hiromasa, S. Wheeler, C. Oppert, S. A. Marshall, J. M. Tomich, and B. Oppert. 2009. Identification, RNAi knockdown, and functional analysis of an ejaculate protein that mediates a postmating, prezygotic phenotype in a cricket. PLoS ONE 4:e7537.

Merle, J. 1968. Ovarian functioning and sexual receptiveness of *Drosophila melanogaster* after implantation of male genital fragments. J. Insect Physiol. 14:1159-1168.

Messina, F. J., A. J. Bloxham, and A. J. Seargent. 2007. Mating compatibility between geographic populations of the seed beetle *Callosobruchus maculatus*. J. Insect Behav. 20:489-501.

Monsma, S. A., H. A. Harada, and M. F. Wolfner. 1990. Synthesis of two *Drosophila* male accessory gland proteins and their fate after transfer to the female during mating. Dev. Biol. 142:465-475.

Murtaugh, M. P., and D. L. Denlinger. 1985. Physiological regulation of long-term oviposition in the house cricket, *Acheta domesticus*. J. Insect Physiol. 31:611-617.

Murtaugh, M. P., and D. L. Denlinger. 1987. Regulation of long-term oviposition in the house cricket, *Acheta domesticus*: roles of prostaglandin and factors associated with sperm. Arch. Insect Biochem. Physiol. 6:59-72.

Neubaum, D. M., and M. F. Wolfner. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:845-857.

Ohashi, Y. Y., K. Hainofukushima, and Y. Fuyama. 1991. Purification and characterization of an ovulation stimulating substance from the male accessory glands of *Drosophila suzukii*. Insect Biochem. 21:413-419.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends Ecol. Evol. 16:364-371.

Peng, J., S. Chen, S. Büsser, H. Liu, T. Honegger, and E. Kubli. 2005. Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. Curr. Biol. 15:207-213.

Pickford, R., A. B. Ewen, and C. Gillott. 1969. Male accessory gland substance: an egg-laying stimulant in *Melanoplus sanguinipes* (F)(Orthoptera, Acrididae). Can. J. Zool. 47:1199-1203.

Price, C. S. C., C. H. Kim, C. J. Gronlund, and J. A. Coyne. 2001. Cryptic reproductive isolation in the *Drosophila simulans* species complex. Evolution 55:81-92.

Quo, F. 1959. Studies on the reproduction of the Oriental migratory locust: The physiological effects of castration and copulation. Acta Ent Sinica 9:464-476.

R Core Development Team 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ram, K. R., and M. F. Wolfner. 2007. Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. PLoS Genet. 3:2428-2438.

Ram, K. R., and M. F. Wolfner. 2009. A network of interactions among seminal proteins underlies the long-term postmating response in *Drosophila*. Proc. Natl. Acad. Sci. U. S. A. 106:15384-15389.

Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone - mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432-449.

Raymond, M., and F. Rousset. 1995. Genepop (version-1.2) - population-genetics software for exact tests and ecumenicism. J. Hered. 86:248-249.

Ritchie, M. G. 2007. Sexual selection and speciation. Annu. Rev. Ecol. Syst. 38:79-102.

Rodriguez-Munoz, R., A. Bretman, J. Slate, C. A. Walling, and T. Tregenza. 2010. Natural and sexual selection in a wild insect population. Science 328:1269-1272.

Ross, C. L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56:2296-2312.

Ross, C. L., and R. G. Harrison. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115:53-68.

Sakaluk, S. K. 1987. Reproductive-behavior of the decorated cricket, *Gryllodes supplicans* (Orthoptera, Gryllidae)- calling schedules, spatial-distribution, and mating Behaviour 100:202-225.

Sakaluk, S. K., J. M. Schaus, A. K. Eggert, W. A. Snedden, and P. L. Brady. 2002. Polyandry and fitness of offspring reared under varying nutritional stress in decorated crickets. Evolution 56:1999-2007.

Schaus, J. M., and S. K. Sakaluk. 2002. Repeatability of sperm number across multiple matings in three cricket species, *Gryllodes sigillatus*, *Gryllus veletis*, and *Gryllus texensis* (Orthoptera: Gryllidae). Can. J. Zool. 80:582-585.

Shapiro, L. H. 2000. Reproductive costs to heterospecific mating between two hybridizing katydids (Orthoptera : Tettigoniidae). Ann. Entomol. Soc. Am. 93:440-446.

Simmons, L. W. 1988. Male size, mating potential and lifetime reproductive success in the field cricket, *Gryllus bimaculatus* (Degeer). Anim. Behav. 36:372-379.

Snook, R., T. Chapman, P. Moore, N. Wedell, and H. Crudgington. 2009. Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. Evol. Ecol. 23:71-91.

Stanley, D., and Y. Kim. 2011. Prostaglandins and their receptors in insect biology. Front Endocrinol 2:1-11.

Stanley-Samuelson, D. W., and J. J. Peloquin. 1986. Egg-laying in response to prostaglandid injections in the Australian field cricket, *Telelgryllus commodus*. Physiol. Entomol. 11:213-219.

Swanson, W. J., and V. D. Vacquier. 2002. Reproductive protein evolution. Annu. Rev. Ecol. Syst. 33:161-179.

Thailayil, J., K. Magnusson, H. C. J. Godfray, A. Crisanti, and F. Catteruccia. 2011. Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. Proc. Natl. Acad. Sci. U. S. A. 108:13677-13681.

Therneau, T. 2011. survival: Survival analysis, including penalised likelihood. R Foundation for Statistical Computing, Vienna, Austria.

Thibout, E. 1979. Stimulation of reproductive activity of females of *Acrolepiosis assectella* (Lepidoptera, Hyponomeutoidea) by the presence of eupyrene spermatozoa in the spermatheca. Entomol. Exp. Appl. 26:279-290.

Tregenza, T., and N. Wedell. 1998. Benefits of multiple mates in the cricket *Gryllus bimaculatus*. Evolution 52:1726-1730.

Turner, L. M., and H. E. Hoekstra. 2008. Causes and consequences of the evolution of reproductive proteins. Int. J. Dev. Biol. 52:769-780.

Wade, M. J., H. Patterson, N. W. Chang, and N. A. Johnson. 1994. Postcopulatory, prezygotic isolation in flour beetles. Heredity 72:163-167.

Wagner, W. E., and C. J. Harper. 2003. Female life span and fertility are increased by the ejaculates of preferred males. Evolution 57:2054-2066.

Wagner, W. E., R. J. Kelley, K. R. Tucker, and C. J. Harper. 2001. Females receive a life-span benefit from male ejaculates in a field cricket. Evolution 55:994-1001.

Warnes, G. R. 2010. gplots: Various R programming tools for plotting data. R Foundation for Statistical Computing, Vienna, Austria.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58:155-183.

Willett, C. S., M. J. Ford, and R. G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79:484-494.

Wolfner, M. F. 1997. Tokens of love: functions and regulation of *Drosophila* male accessory gland products. Insect Biochem. Mol. Biol. 27:179-192.

Wolfner, M. F. 2009. Battle and ballet: molecular interactions between the sexes in *Drosophila*. J. Hered. 100:399-410.

Xu, J., and Q. Wang. 2011. Seminal fluid reduces female longevity and stimulates egg production and sperm trigger oviposition in a moth. J. Insect Physiol. 57:385-390.

Xue, L., and M. Noll. 2000. *Drosophila* female sexual behavior induced by sterile males showing copulation complementation. Proc. Natl. Acad. Sci. U. S. A. 97:3272-3275.

Yamane, T., and T. Miyatake. 2010. Induction of oviposition by injection of male-derived extracts in two *Callosobruchus* species. J. Insect Physiol. 56:1783-1788.

Yi, S. X., and C. Gillott. 1999. Purification and characterization of an oviposition-stimulating protein of the long hyaline tubules in the male migratory grasshopper, *Melanoplus sanguinipes*. J. Insect Physiol. 45:143-150.

Zuk, M., and L. Simmons. 1997. Reproductive strategies of the crickets (Orthoptera: Gryllidae). Pp. 89-109 *in* J. Choe, and B. Crespi, eds. The Evolution of Mating Systems in Insects and Arachnids. Cambridge University Press, Cambridge.

-

CHAPTER 3

STRUCTURE OF A MOSAIC HYBRID ZONE BETWEEN THE FIELD CRICKETS *GRYLLUS FIRMUS* AND *G. PENNSYLVANICUS*

Abstract

Hybrid zones provide insight into the nature of species boundaries and the evolution of barriers to gene exchange. Characterizing multiple regions within hybrid zones is essential for understanding both their history and current dynamics. Here, we describe a previously uncharacterized region of a well-studied hybrid zone between two species of field crickets, *Gryllus pennsylvanicus* and *G. firmus*. We use a combination of mitochondrial DNA sequencing, morphological data and modeling of environmental variables to identify the ecological factors structuring the hybrid zone and define patterns of hybridization and introgression. We find an association between species distribution and natural habitat; *G. pennsylvanicus* occupies natural habitat along forest edges and natural clearings, while *G. firmus* occupies more disturbed areas in agricultural and suburban environments. Hybridization and introgression occur across patch boundaries; there is evidence of substantial admixture both in morphological characters and mtDNA, over a broad geographic area. Nonetheless, the distribution of morphological types is bimodal. Given that F1 hybrids are viable and fertile in the lab, this suggests that strong prezygotic barriers are operating in this portion of the hybrid zone.

Introduction

Hybrid zones have been described as 'natural laboratories for evolutionary studies' (Hewitt 1988; Barton and Hewitt 1989) and 'windows on evolutionary process' (Harrison 1990, 1993). They are places where diverged lineages meet and interact, providing insight into the genetic architecture of speciation and the evolutionary forces that shape divergence (Barton and Hewitt 1985; Harrison 1990; Payseur 2010). Hybrid zone studies have demonstrated that species boundaries are semipermeable and that permeability varies across the genome (Key 1968; Barton

and Hewitt 1981; Harrison 1986, 1990; Wu 2001). Genomic regions that contain alleles contributing to reproductive isolation will either prevent the formation of hybrids or decrease hybrid viability or fertility, restricting introgression. Recombination over multiple generations breaks up parental genomes and some segments of the genome can then be freely exchanged between species. From studies of differential introgression we can identify genomic regions that are under selection, estimate the strength of that selection, and ultimately link some of those regions to reproductive barriers (Harrison 1990; Payseur 2010).

Many hybrid zones may be "tension zones," in which parental types persist because of selection against hybrid genotypes, independent of environment (Barton and Hewitt 1985). However, hybrid zones may also be maintained by environmental selection favoring different parental forms in different ecological settings. In the latter case, hybrid zones can be clinal, with parental forms favored on either side of an ecotone and hybridization occurring in the center (Endler 1977). Alternatively, hybrid zones can be mosaic, with parental forms patchily distributed across heterogeneous habitat and hybridization occurring across patch boundaries or in intermediate habitats (Harrison 1986; Harrison and Rand 1989; Rand and Harrison 1989; Ross and Harrison 2002), but see Searle (1993). In a heterogeneous landscape, hybrid zones may exhibit a mix of different dynamics (e.g. *Bombina* hybrid zone; Szymura and Barton 1991; Vines et al. 2003; Yanchukov et al. 2006), and as a consequence, reproductive barriers and patterns of introgression may vary geographically (Teeter et al. 2010). Therefore, characterizing multiple transects or regions is essential for understanding both hybrid zone history and hybrid zone dynamics.

Here, we describe a previously uncharacterized region of a well-studied hybrid zone between two species of North American field crickets, *Gryllus pennsylvanicus* and *G. firmus*.

The hybrid zone has been carefully characterized in Virginia and Connecticut, and reproductive barriers are known to vary between these two regions. This paper examines patterns of variation in Pennsylvania, compares these patterns with those seen elsewhere, and investigates what ecological factors maintain the structure of the hybrid zone.

Field cricket hybrid zone

The hybrid zone between *G. pennsylvanicus* and *G. firmus* stretches at least from southern Connecticut to Virginia along the eastern slopes of the Appalachian, Blue Ridge and Northern Highland Mountains (Harrison and Arnold 1982). The glacial history of the northeastern United States and the distribution of *Gryllus* mitochondrial DNA (mtDNA) haplotypes provide strong evidence that the hybrid zone formed as a result of secondary contact between lineages that diverged in allopatry (Harrison et al. 1987; Willett et al. 1997; Maroja et al. 2009a). *Gryllus pennsylvanicus* extends west from the Appalachian and Blue Ridge Mountains and through the mountains to the south. *Gryllus firmus*, also known as the beach cricket, occurs to the east of the Appalachian Mountains throughout the Piedmont, coastal plain and along beaches south into Florida (Alexander 1957, 1968; Harrison and Arnold 1982). Both species occupy grassy, disturbed habitats and can be found under rocks, debris or clumps of vegetation. Both species are univoltine in the north; females lay eggs in the soil, eggs diapause over the winter, hatch in the spring and adults emerge in late summer or early fall (Fulton 1952). In the south *G. firmus* is multivoltine and females lay both diapause and non-diapause eggs (Fulton 1952; Alexander 1968; Walker 1980).

The two cricket species diverged about 200,000 ya (Willett et al. 1997; Maroja et al. 2009a). They are very similar morphologically and were considered part of a single variable

species until the 1950s (Fulton 1952; Alexander 1957). *Gryllus pennsylvanicus* is, on average, a smaller cricket, with darker tegmina (modified leathery front wing) and a relatively shorter ovipositor (Figure 3.1). Calling songs of the two species (used to attract females at long distances) are very similar but have slightly different pulse and chirp rates; courtship songs (used to initiate mating) are identical (Alexander 1957; Doherty and Storz 1992). Despite these similarities, there is evidence of behavioral isolation, with females of both species reluctant to mate with heterospecific males in the laboratory (Maroja et al. 2009b). The crickets are also isolated by post-mating pre-zygotic barriers in one direction; *G. firmus* females mated with *G. pennsylvanicus* males lay few eggs, none of which are fertilized (Harrison 1983; Maroja et al. 2009b; Larson et al. 2012).

In Virginia the two species are temporally isolated because of differences in development time, with *G. firmus* adults emerging later in the fall (Harrison and Arnold 1982; Harrison 1985). In Connecticut, adults emerge synchronously, but are associated with different soil types; *G. firmus* on sandy soils and *G. pennsylvanicus* on loamy soils (Harrison 1986; Rand and Harrison 1989; Ross and Harrison 2002). What maintains this soil association remains unclear; females of both species prefer to oviposit in loamy soils in the laboratory (Ross 2000) and the viability of overwintering diapause eggs appears to be independent of soil type (Ross and Harrison 2006).

Although multiple barriers isolate these species and very few F1 hybrids are present within the hybrid zone, evidence for introgression is clear (Harrison and Bogdanowicz 1997; Ross and Harrison 2002). Some barriers appear to be consistent across different regions of the hybrid zone (assortative mating, one-way fertilization incompatibility), while other barriers vary from one ecological setting to another (development time, soil association). Documenting patterns of species distributions and their ecological context builds the foundation for comparing

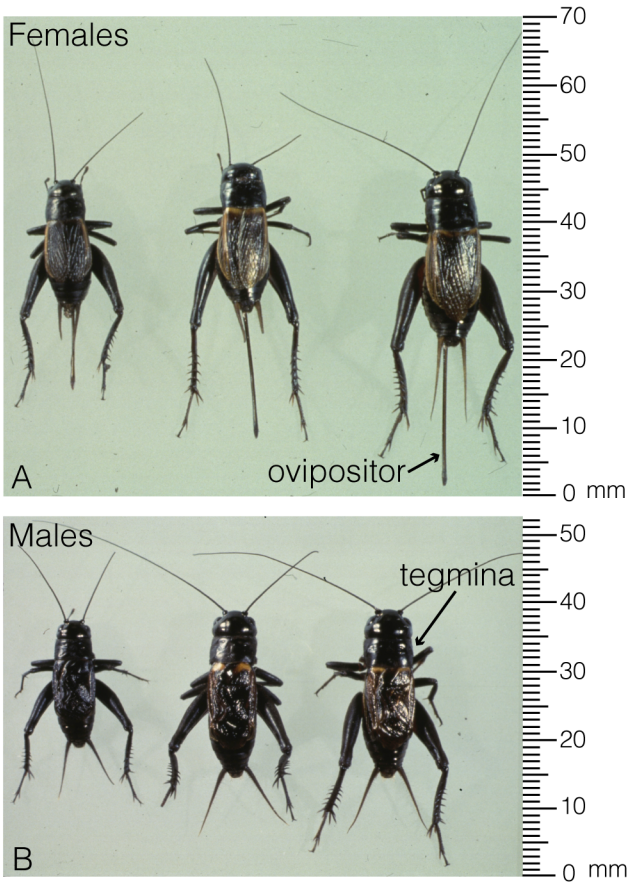


Figure 3.1 Morphological variation in *G. pennsylvanicus* and *G. firmus*. Representative crickets of *G. pennsylvanicus* (left), *G. firmus* (right) and intermediate individuals (middle): females (above) and males (below). Arrows indicate the female ovipositor and male tegmen. *Gryllus firmus* are typically larger, with relatively longer ovipositors and lighter tegmina.

patterns of introgression across different regions of the hybrid zone. Here, we describe patterns of variation in a previously uncharacterized region of the hybrid zone in south-central Pennsylvania.

Materials and Methods

Cricket Sampling

In the fall of 2010, we collected 104 crickets from 9 localities (EST, PTJ, GUI, TRI, MAY, MOT, ELK, POW, ITH) in the northeastern range of *G. firmus* and G. *pennsylvanicus*. To these, we added an additional 26 crickets from 4 localities collected by Maroja et al. (2009a) (SCR, SCO, MET, MOO). We also used mtDNA sequence data for 98 crickets from 28 localities described in Willet et al. (1997) and Maroja et al. (2009a) (Table S3.1). These samples provide a broad geographic context for analyzing the distribution of cricket mitochondrial DNA (mtDNA) haplotypes.

In late summer/fall of 2008 and 2010, we collected 877 crickets from 88 localities within a small region of the hybrid zone in south-central Pennsylvania (Table 3.1). Collection localities span the transition from the Appalachian Mountains into the Great Appalachian Valley. In this region, the Appalachians form a series of continuous ridges and intervening valleys that can range in elevation from 100m to 650m (elevation can change from 250m to 570m over only 1.8 km). The ridges are broken by several narrow and dramatic gaps where the Susquehanna River and other small waterways cross the mountains. The Great Appalachian valley is an extended chain of lowlands bounded by the Appalachian Mountains to the west and the Blue Ridge Mountains and Northern Highlands to the east, and includes the Shenandoah Valley in northern Virginia. To the east of the Blue Ridge Mountains are the lowlands of the Piedmont and the

Table 3.1 Collecting localities for crickets in the Pennsylvania portion of the hybrid zone. N = total number of crickets collected at each locality, $N^{\dagger}$ = number of crickets genotyped for mitochondrial DNA. Habitat includes both the type of substrate crickets were collected from under (e.g. rocks, trash, woodpiles, etc.) and the type of surrounding habitat (e.g. forest, field, etc.). Sites in bold contains long-wing morph crickets.

Locality	ID	Species	N	N†	Lat. (N)	Long. (W)	Ele. (m)	Habitat	Date
Echo Valley, PA	A	*Gp*	4	-	40°35'33"	76°24'36"	205	rocks/forest clearing	2008
Outwood, PA	B	*Gp*	2	-	40°31'38"	76°28'24"	154	rocks/wooded roadside	2008
Indiantown Gap, PA	C	*Gp*	8	5	40°26'08"	76°35'54"	156	rocks/forest clearing stream	2008
Quentin, PA	D	*Gp*	19	9	40°15'51"	76°26'08"	223	rocks/wooded roadside	2008
Cornwall, PA	E	*Gp*	3	-	40°16'36"	76°24'39"	172	rocks/edge of cornfield	2008
Schaefferstown, PA	F	*Gp*	5	-	40°17'37'	76°17'52"	178	rocks/grassy field	2008
Womelsdorf, PA	G	admix (*Gp*)	5	-	40°23'24"	76°11'54"	146	trash pile/grassy roadside	2008
Rehrersburg, PA	H	*Gp*	7	7	40°26'57"	76°13'58"	169	burrows/grassy slope	2008
Reading, PA	I	admix	12	12	40°22'51"	76°01'46"	104	rocks/forest clearing	2008
Nottingham, PA	J	admix	10	10	39°44'15"	76°02'48"	111	rocks/forest clearing	2008
Nottingham, PA	K	admix (*Gp*)	3	3	39°44'32"	76°02'02"	134	trash pile/city park	2008
Holtwood, PA	L	admix (*Gp*)	8	8	39°48'49"	76°19'42"	48	rocks/forest clearing	2008
Marysville, PA	M	*Gp*	4	-	40°20'28"	76°54'36"	92	rocks/shoreline	2008
Millersburg, PA	N	admix (*Gp*)	5	5	40°32'07"	76°57'59"	112	rocks/wooded beach	2008
Bloomsburg, PA	O	admix (*Gp*)	4	4	40°58'39"	76°28'10"	142	rocks/wooded campground	2008
Cattawissa, PA	P	admix (*Gp*)	3	-	40°56'55"	76°30'53"	145	rocks/wooded roadside	2008
Shamokin Dam, PA	Q	*Gp*	3	-	40°51'22"	76°48'32"	133	rocks/boat launch	2008
Northumberland, PA	R	*Gp*	19	9	40°53'02"	76°48'16"	202	trash cans/forest clearing	2008
Etters, PA	S	admix (*Gf*)	1	-	40°09'00"	76°44'58"	89	rocks/grassy field in park	2008
York Haven, PA	T	admix (*Gf*)	1	-	40°06'42"	76°42'36"	76	rocks/wooded shoreline	2008
Mt Wolf, PA	U	admix	21	13	40°03'48"	76°42'34"	123	rocks/grassy field slope	2008
York, PA	V	*Gf*	1	1	39°58'59"	76°44'00"	116	trash can/motel parking lot	2008
York, PA	W	admix (*Gf*)	1	1	39°58'01"	76°46'38"	126	trash can/gas station	2008
Mt Royal, PA	X	admix (*Gf*)	4	4	40°02'35"	76°53'45"	115	rocks/grassy roadside slope	2008
Pinchot Lake	Y	*Gp & Gf*	5	5	40°04'06"	76°54'32"	146	rocks/ wooded boat launch	2008
Dillsburg, PA	Z	admix	6	5	40°05'18"	77°01'19"	219	rocks/wooded roadside	2008
Locust Point, PA	AA	*Gp*	3	-	40°11'02"	77°03'19"	151	rocks/grassy slope	2008
Wertzville, PA	AB	*Gp*	3	-	40°16'30"	77°02'46"	140	rocks/wooded roadside	2008

Table 3.1 (Continued)

Goodhope, PA	AC	admix (*Gp*)	5	-	40°17'17"	77°00'11"	135	rocks/wooded roadside	2008
Hummelstown, PA	AD	admix (*Gp*)	3	3	40°15'43"	76°41'18"	128	rocks/edge of cornfield	2008
Harper Tavern, PA	AE	admix (*Gp*)	5	5	40°24'17"	76°34'32"	118	rocks/grassy roadside slope	2008
Jonestown, PA	AF	admix (*Gp*)	10	5	40°24'55"	76°29'51"	141	rocks/grassy roadside slope	2008
Carlisle, PA	AG	admix (*Gp*)	6	1	40°12'21"	77°16'22"	156	rocks/grassy roadside slope	2008
Newville, PA	AH	*Gp*	1	-	40°12'17"	77°24'46"	179	rocks/grassy roadside slope	2008
Newville, PA	AI	admix	5	5	40°08'21"	77°21'56"	187	rocks/pasture	2008
Michaux Forest	AJ	*Gp*	7	5	40°03'26"	77°17'46"	364	rocks/wooded roadside	2008
Michaux Forest	AK	*Gp*	7	6	39°58'08"	77°22'46"	500	rocks/wooded roadside	2008
Gettysburg, PA	AL	admix (*Gf*)	6	4	39°52'40"	77°14'38"	198	rocks/grassy roadside slope	2008
Emmitsburg, MA	AM	admix	9	7	39°42'11"	77°19'00"	125	trash/freeway onramp	2008
Carroll Valley, PA	AN	*Gp*	5	4	39°44'18"	77°23'33"	180	rocks/wooded roadside	2008
Rouzerville, PA	AO	admix	5	3	39°44'17"	77°31'11"	557	trash/parking lot	2008
Waynesboro, PA	AP	admix (*Gf*)	2	1	39°45'47"	77°35'52"	195	rocks/grassy roadside slope	2008
Mercersburg, PA	AQ	admix	7	5	39°49'18"	77°53'56"	190	rocks/cemetery lawn	2008
Charlestown, PA	AR	*Gp & Gf*	5	5	39°52'15"	77°57'12"	269	rocks/forest clearing	2008
Fort Loudon, PA	AS	admix (*Gf*)	5	5	39°54'41"	77°54'19"	186	rocks/wooded roadside	2008
Saint Thomas, PA	AT	admix (*Gf*)	1	1	39°54'24"	77°50'24"	212	rocks/parking lot	2008
Saint Thomas, PA	**AU**	admix	20	5	39°53'47"	77°47'02"	164	concrete blocks/grassy field	2010
Saint Thomas, PA	**AV**	admix	21	-	39°53'06"	77°49'07"	196	concrete blocks/churchyard	2010
Fort Loudon, PA	**AW**	admix	16	15	39°54'49"	77°54'19"	192	trash cans/city park	2010
Big Mountain	AX	admix	25	14	39°55'44"	77°57'18"	698	rocks/forest clearing	2010
Fayetteville, PA	**AY**	admix	24	-	39°55'09"	77°33'18"	237	woodpile/grassy field	2010
South Mountain, PA	**AZ**	admix	27	15	39°50'37"	77°28'40"	510	concrete blocks/churchyard	2010
Quincy, PA	BA	admix (*Gf*)	9	5	39°48'38"	77°34'08"	235	concrete blocks/churchyard	2010
Five Forks, PA	BB	admix	14	5	39°47'56"	77°36'37"	228	boards/churchyard	2010
Milnor, PA	**BC**	admix	27	-	39°45'48"	77°46'04"	179	boards/city park	2010
Bino, PA	**BD**	admix (*Gf*)	10	-	39°45'57"	77°47'49"	171	woodpile/churchyard	2010
Shimpsown, PA	BE	*Gp & Gf*	2	-	39°47'17"	77°50'34"	147	trash pile/grassy field	2010

Table 3.1 (Continued)

Big Mountain	BF	*Gp*	3	-	39°59'25"	77°55'51"	382	rocks/forest clearing	2010
Scotland, PA	**BG**	admix	30	5	39°56"51"	77°33'41"	258	woodpile/churchyard	2010
Fayetteville, PA	**BH**	admix (*Gp*)	10	-	39°54'21"	77°31'51"	264	rocks/wooded roadside	2010
Cashtown, PA	BI	admix (*Gp*)	9	-	39°53'07"	77°22'04"	267	rocks/wooded roadside	2010
Michaux Forest	**BJ**	admix (*Gp*)	5	-	39°51'36"	77°26'18"	482	rocks/wooded roadside	2010
Charmian, PA	BK	*Gp & Gf*	12	12	39°44'22"	77°28'11"	416	boards/churchyard	2010
Chambersburg, PA	**BL**	admix	28	-	39°54'46"	77°42'16"	229	trashcans/city park	2010
New Franklin, PA	BM	admix	11	-	39°52'48"	77°38'18"	219	woodpile/churchyard	2010
New Franklin, PA	BN	admix	12	5	39°52'07"	77°38'09"	248	boards/churchyard	2010
Pond Bank, PA	**BO**	admix	23	5	39°52'21"	77°32'36"	272	rocks/churchyard	2010
Kauffman, PA	BP	admix	26	5	39°50'04"	77°42'01"	195	boards/city park	2010
Cashtown, PA	**BQ**	admix	32	5	39°52'03"	77°43'41"	201	concrete blocks/churchyard	2010
Saint Thomas, PA	**BR**	admix	8	-	39°53'59"	77°45'32"	197	boards/churchyard	2010
Markes, PA	BS	admix (*Gf*)	4	-	39°52'29"	77°52'20"	172	boards/city park	2010
Sylvan, PA	**BT**	*Gp & Gf*	25	-	39°45'16"	78°01'27"	154	boards/lumberyard	2010
Lemasters, PA	BU	admix	18	16	39°51'23"	77°51'41"	174	trash cans/city park	2010
Williamson, PA	BV	admix	20	15	39°51'12"	77°48'04"	165	trash cans/city park	2010
Chambersburg, PA	**BW**	admix (*Gf*)	10	-	39°57'41"	77°43'48"	170	trash cans/city park	2010
Edenville, PA	**BX**	admix	22	5	39°57'35"	77°47'54"	219	boards/churchyard	2010
Charlestown, PA	BY	*Gp*	2	-	39°52'10"	77°57'23"	339	rocks/forest clearing	2010
Harrisonville, PA	BZ	admix (*Gp*)	5	5	39°59'16"	78°03'45"	241	concrete blocks/churchyard	2010
Bedford, PA	CA	admix (*Gp*)	4	4	40°05'03"	78°31'31"	361	trash cans/city park	2010
Osterburg, PA	CB	admix (*Gp*)	14	5	40°10'25"	78°31'42"	354	grass clumps/grassy field	2010
PA Turnpike	CC	*Gf*	3	-	40°01'05"	78°11'23"	538	woodpile/grassy roadside	2010
Saluvia, PA	CD	admix	10	9	39°59'05"	78°06'36"	333	grass clumps/grassy field	2010
Andover, PA	CE	admix	15	15	39°55'43"	78°06'28"	306	trash cans/churchyard	2010
Needmore, PA	CF	admix (*Gp*)	7	7	39°49'40"	78°15'15"	273	burrows/roadside slope	2010
Breezewood, PA	**CG**	admix (*Gf*)	17	5	39°59'58"	78°14'14"	391	concrete blocks/hotel lawn	2010
Everett, PA	**CH**	admix	7	6	40°00'31"	78°22'35"	322	burrows/bank along river	2010

Table 3.1 (Continued)

Hopewell, PA	CI	admix	17	16	40°08'29"	78°19'59"	318	burrows/roadside slope	2010
Roaring Spring, PA	CJ	admix (*Gp*)	10	5	40°20'27"	78°52'50"	417	rocks/grassy roadside slope	2010
		Total	**878**	**375**					

coastal beaches. There is a large gap in the eastern ridge of mountains between the Reading Prong (near collecting locality I) and South Mountain (collecting locality AJ) through which the Susquehanna River passes, connecting the Great Appalachian Valley with the Piedmont region.

In general, the mountains are heavily forested (poor cricket habitat), but have some natural clearings and are dissected by roadways and water gaps. The mountain valleys are typically moderately populated farmland, while the relevant portion of the Great Appalachian valley (Lehigh, Lebanon and Cumberland valleys) is relatively heavily populated, and primarily agricultural or suburban.

Live crickets were brought back to laboratory and frozen at -80^{o}C. The majority of crickets were collected as adults, but in some cases, crickets were collected as late instar nymphs. Nymphs were allowed to mature in the laboratory before freezing.

Mitochondrial DNA

We sequenced the mtDNA gene cytochrome oxidase I (COI) for a total of 130 crickets from 13 localities across the hybrid zone and 119 crickets from 31 localities within the Pennsylvania hybrid zone. We isolated genomic DNA from a single femur using the DNeasy tissue kit (QIAGEN, Valencia, CA). Locus specific primers were used to amplify a 1.9 kb fragment of the mitochondrial DNA gene cytochrome oxidase I (COI) and the adjacent tRNA: *G. veletis* COI F (102) (5′- ACCCCCATCATTAACCCTTTTA- 3′)(Maroja et al. 2009a) and Eva/3372 (1885) (5′- GAGACCATTACTTGCTTTCAGTCATCT - 3′) (Simon et al. 1994), and a set of internal primers designed from *G. firmus* mtDNA sequence were used for samples with shorter sequence length: cricketCOI.595 (5′- ATTTACGGTTGGAATAGATGTTGATACCC - 3′) and cricketCOI.1270 (5′- GAAGCTTAAATTCATCGCACTTTTCTG- 3′). DNA was

amplified using polymerase chain reaction (PCR) in 10 μL reactions containing: 1 μL genomic DNA, 2mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μM of each forward and reverse primer and 0.1 μL (0.5 U) Platinum Taq polymerase (Invitrogen, San Diego, CA) in 1x PCR buffer (20 mM Tris-HCL, pH 8.4, 50 mM KCl). PCR was performed using an initial denaturation of 95°C for 2 min followed by a touchdown protocol of 35 cycles of 95°C for 50 s, 65-55°C for 1 min (the annealing temperature decreased by 1°C each cycle for the first 10 cycles) and 72°C for 1 min. Samples were enzymatically cleaned with EXOSAP, sequenced in both directions with Big Dye chemistry and analyzed on an ABI 3730 automated sequencer at the Cornell University Life Sciences Core Laboratories Center for Genomics.

The resulting chromatograms were base called using the phred-phrap algorithm and assembled in CodonCode Aligner software (CodonCode Corp, Dedham, MA). All assembled sequences were trimmed and visually inspected. We included an additional 28 sequences from Willet et al. (1997) and 70 sequences from Maroja et al. (2009a) of mtDNA COI and constructed a phylogeny using MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001). To select the optimal substitution model we used hierarchical likelihood ratio tests implemented in jModeltest version 0.1.1 (Posada 2008). To generate trees we used the general time reversible model with invariant sites, gamma rates, and default priors (GTR + I + G), allowing the rate at each site to change over evolutionary history. We ran searches for ten million generations, sampling every 2,000 generations and discarded trees from the first 4,000,000 generations (burn-in time). We constructed a 50% majority-rule consensus tree from the remaining trees. The phylogenetic tree was rooted using three *G. rubens* sequences from Maroja et al. (2009a). MrBayes was run using the resources of the Cornell University Computational Biology Service Unit.

We used the Sequenom MassARRAY platform to genotype crickets for twelve mtDNA

single nucleotide polymorphisms (SNPs). We collected 301 crickets across 46 sampling localities in Pennsylvania and 81 crickets from 7 localities across the hybrid zone. A total of 66 crickets had previously been sequenced for the entire mtDNA COI gene (Table 3.1), and were used to validate our genotyping results. We assayed five SNPs in the mtDNA COI gene (site numbers 796, 952, 1036, 1204, 1382 from Willet et al. 1997) and seven mtDNA SNPs identified in Andres et al. (2013). Multiplexed site-specific primers were used to amplify target DNA, followed by a single base extension of a primer immediately adjacent to the target SNP. The resulting product was analyzed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) and the mass difference of each possible SNP nucleotide allowed unambiguous genotyping. Assays were designed using the MassARRAY Assay Design version 4.0 (Sequenom, San Diego, CA, USA). Amplicon primer sequences, amplicon length, annealing temperature, extend primer sequence and target SNP for each assay are listed in Table S3.2. Reactions were performed using iPLEX Gold chemistry at the Cornell Life Sciences Core Laboratories Center for Genomics and SNP genotypes were called using the Sequenom MassARRAY Typer Analysis version 4.0. We constructed a phylogeny of the mtDNA SNPs using MrBayes as described above (generations: 10,000,000; sampling: 2,000; burn-in: 4,000,000; 50% majority-rule consensus).

Morphological measurements

We characterized morphological traits that distinguish *G. firmus* and *G. pennsylvanicus* for crickets from nine allopatric populations (*G. firmus*: GUI, MAY, MOT, TOM, MET; *G. pennsylvanicus*: ITH, NBL, SCR, SCO) and from all collecting localities in our focal study area in Pennsylvania. We measured three morphological characters for each cricket: body length

(BL), femur length (FL), and pronotal width (PW). In addition, we measured the tegmina color (TEG) in male crickets and the ovipositor length (OL) in female crickets (Figure 3.1). Body length, femur length and pronotum width all reflect overall size differences (*G. firmus* is typically larger than *G. pennsylvanicus*). Tegmina color (males) is lighter and ovipositor length (females) is greater in *G. firmus*. Ovipositor length is the character that most clearly differentiates the two species (Harrison and Arnold 1982; Harrison 1986). We also recorded the presence/absence of fully developed long hind-wings on both males and females. All size measurements were made to the nearest 0.1 mm using a pair of vernier calipers.

To measure the color of the male tegmina, we used a USB4000 spectrophotometer with a PX-2 pulsed xenon lamp (Ocean Optics, Dunedin, FL) to capture spectral reflectance. The probe was mounted in a metal stand at a 90^{o} angle 0.7 mm from the surface of the tegmina. For each male, we recorded and averaged spectral reflectance for three points near the center of the tegmina. We used the program SpectraSuite version 2.0 (Ocean Optics) to capture the wavelength readings. All spectral measurements were recorded as the percentage of reflected light relative to a Spectralon white standard (Ocean Optics). We restricted our analyses to wavelengths of 300 – 700 nm and used a segmental classification method to quantify three aspects of color: brightness, chroma, and hue (Endler 1990) using the software program CLR (Montgomery 2008). We calculated total brightness (B) as $R_{300\text{-}700}$, the summed reflectance from 300 nm to 700 nm. We also divided our reflectance data into four bins of 100 nm each, calculated the total brightness for each bin (B_r = 600 -700, B_y = 500 – 600, B_g = 400 – 500, and B_b = 300 – 400) and then calculated chroma: $\sqrt{(B_r - B_g)^2 + (B_y - B_b)^2}$ and hue: $\arctan[(B_y\text{-}B_b)/B]/[(B_r\text{-}B_g)/B]$.

Analysis of morphological data

We used principal component analysis (PCA) to explore variation in morphological data. We performed separate PCAs for male body size (BL, FL, PW), tegmina color (brightness, chroma and hue) and female body size (BL, FL, PW) using singular value decomposition of the scaled and centered morphological data with the function 'prcomp' in R (R Core Development Team 2010). We performed a one-way analysis of variance (ANOVA) for each morphological trait (ovipositor length and the first principal components of male body size, male tegmina color, and female body size) to test for differences between allopatric *G. firmus* and *G. pennsylvanicus* populations. We used a linear discriminant analysis to determine how well each of these morphological traits classifies allopatric crickets. ANOVA and linear discriminant analyses were performed in R using the packages '*stats*' (R Core Development Team 2010) and '*MASS*' (Venables and Ripley 2002).

To identify morphological clusters and estimate cluster membership we used a fuzzy c-means clustering algorithm using the 'fanny' function from the R package '*cluster'* (Maechler et al. 2002). In contrast to hard clustering algorithms (such as K-means) in which data elements are divided into distinct clusters, fuzzy clustering allows data elements to belong to more than one cluster and assigns a corresponding set of membership levels (or membership coefficients). Because we were interested in identifying morphological clusters for the two parental species and admixed individuals, we used $k = 2$, which divided individuals into 2 morphological clusters and a third "fuzzy" or admixed cluster. The membership exponent (r) determines the degree of 'fuzziness' in cluster assignment. A value of $r = 1.0$ assigns each data element to a single cluster and is equivalent to a classic K-means clustering, while values of $r > 1.0$ become increasingly fuzzy until all individuals are equally distributed among the k clusters (i.e. all belong to single

"fuzzy" cluster). The proportion of individuals classified as admixed depends on the choice of *r*, but there is no clear rule for selecting *r* values. We used an approach, similar to other ecological studies (Schaefer and Wilson 2002; Gompert et al. 2010), of conducting separate fuzzy c-means clustering analyses for male morphology (male body size and tegmina color) and female morphology (female body size and ovipositor length) using *r* values ranging from 1.0 to 2.5. Values of $r > 1.75$ for male traits and > 2.0 for female traits classified nearly all individuals as fuzzy. Therefore, we report cluster assignments using three values of *r* for male traits (1.25, 1.5, 1.75) and three values of *r* for female traits (1.5, 1.75, 2.0), all using $k = 2$ (Table 3.2). Values of $r = 1.25$ for males and $r = 1.75$ for females delineates cluster membership in a manner consistent with other clustering algorithms and morphological indices that have been applied to these crickets (Harrison and Arnold 1982; Harrison 1986; Rand and Harrison 1989; Harrison and Bogdanowicz 1997) and were used for all further analyses.

Environmental predictors of species distributions

We assessed twelve environmental variables for each of our 88 sampling sites at a 1 km scale for all variables. We calculated percent natural vegetation cover based on a 30 m resolution land cover raster from circa 2005 (Homer et al. 2007). We considered urban, pasture, agriculture, silviculture, and recreational (e.g., golf-courses) land-cover types as non-natural. We calculated terrain complexity using the raster calculator feature in ArcGIS 9.3, where each elevation pixel was assigned the variance of the neighbor pixels (Huaxing 2008). This metric provides significant information on habitat heterogeneity and microclimate turnover. We assessed physical soil characteristics for each sampling location (i.e., maximum % sand, silt, clay, and organic matter) using spatial data made available by Soil Data Mart (Soil Survey Staff 2012). We

Table 3.2 Morphological clustering. Classification of crickets from the Pennsylvania hybrid zone as *G. firmus*, *G. pennsylvanicus* and admixed based on fuzzy c-means clustering of morphological characters with k = 2.

Character	Cluster	$r = 1.25$	$r = 1.50$	$r = 1.75$	$r = 2.0$
Males	*Gp*	**43.4%**	37.0%	28.5%	-
(N = 424)	*Gf*	**38.7%**	33.7%	23.3%	-
	Mixed	**17.9%**	29.2%	48.1%	-
Females	*Gp*	-	**37.9%**	42.3%	32.8%
(N = 420)	*Gf*	-	**48.6%**	33.0%	25.6%
	Mixed	-	**13.6%**	24.7%	41.6%

also recorded vegetation density (USGS & FAO 2000), latitude, elevation (Jarvis et al. 2009), human footprint (Sanderson et al. 2002), annual temperature (Bio 1; Hijmans et al. 2005), and annual precipitation (Bio 12; Hijmans et al. 2005).

We analyzed our data using simple linear regressions (standard least squares). We used this univariate approach to test the relationship of each explanatory variable with ovipositor length or morphological clustering membership coefficient. We then used model selection tests including all environmental variables and their interactions to find the combinations of variables that best explained ovipositor length and cluster membership. Competing models were ranked based on Akaike Information Criterion (AICc), and we reported the model with the highest goodness-of-fit for each run. Linear regression and automated model selection were conducted using JMP 10.0 (SAS 2012).

Results

Mitochondrial DNA

Phylogenetic analysis of the mtDNA COI gene produced a tree with four major groups, each group composed of conspecific crickets found primarily in circumscribed geographical areas: (1) northern *G. pennsylvanicus*, (2) southern *G. pennsylvanicus*, (3) northern *G. firmus*, and (4) southern *G. firmus* (Figure S3.1). These four groups correspond to the haplotype groups identified by Willett et al. (1997) and Maroja et al. (2009a). Analysis of seven mtDNA SNPs identified the same four major haplotype groups, and two of these SNPs (Table S3.2, SNPs 448 and 554) were shared among the majority of *G. firmus* crickets (Figure 3.2A). In both mtDNA phylogenies, we color each group (yellow: northern *G. pennsylvanicus*; orange: southern *G.*

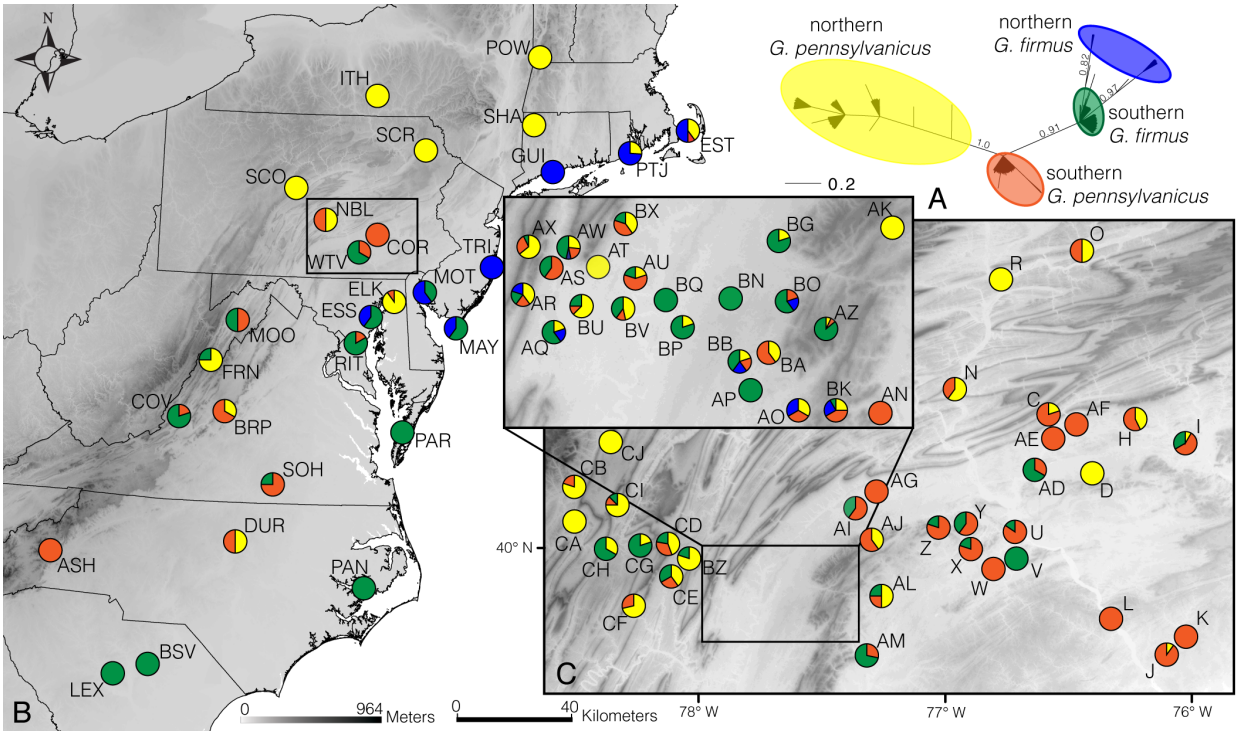


Figure 3.2 Mitochondrial DNA phylogeny and haplotype distribution for *G. firmus* and *G. pennsylvanicus*. Yellow and orange represent northern and southern *G. pennsylvanicus* haplotypes and blue and green represent northern and southern *G. firmus* haplotypes. Each pie shows the proportion of crickets belonging to a mtDNA group and letters refer to the location details in Tables 3.1 and S3.1. **A)** Bayesian 50% majority-rule consensus tree of five mtDNA SNPs from cytochrome oxidase I (site numbers 796, 952, 1036, 1204, 1382 from Willet et al. 1997) and seven mtDNA SNPs identified in Andres et al. (2013) (Table S3.2). Values on the branches correspond to the Bayesian posterior probabilities. The tree includes 81 crickets from 7 localities across the hybrid zone (Table S3.1) and 301 crickets from 46 sampling localities within the Pennsylvania hybrid zone. **B)** Distribution of mtDNA haplotypes across the hybrid zone. The rectangle highlights the location of the Pennsylvania study area. Shading indicates elevation (m) with higher elevations represented by darker shades of gray. **C)** Detailed map of mtDNA haplotypes within Pennsylvania.

pennsylvanicus; blue: northern *G. firmus* and green: southern *G. firmus*) and use these colors to represent the proportion of crickets belonging to each mtDNA group for each sampling locality across the eastern U.S. (Figure 3.2B) and across our focal study area in Pennsylvania (Figure 3.2C).

Morphological data

The average size measurements for crickets from allopatric populations and collecting localities in the Pennsylvania hybrid zone are listed in Table S3.3 and Table S3.4. The first principal component of male body size explained the majority of variation in body size among male crickets (87.7% ± 1.62) and all three measurements of male body size (body length, femur length and pronotum width) had positive loadings on PC1. We used PC1 to represent male body size for all further analyses and we refer to this component as 'male body size'. We found significant differences in male body size between *G. firmus* (N = 72, -0.19 ± 1.32) and *G. pennsylvanicus* (N = 63, -1.51 ± 1.22) (ANOVA: $F_{1,133} = 35.94$, $P < 0.0001$), with *G. firmus* having larger body sizes (Figure 3.3A). The first principal component of male tegmina color explained most of the variation in male tegmina color (75.0% ± 1.50) and all three measurements (tegmina brightness, chroma and hue) had positive loadings on PC1. We refer to this component as 'tegmina color'. Tegmina color differed significantly between *G. firmus* (N = 72, 0.42 ± 1.41) and *G. pennsylvanicus* (N = 63, -1.53 ± 1.04) (ANOVA: $F_{1,133} = 81.48$, $P < 0.0001$) with *G. firmus* having lighter tegmina (Figure 3.3B). The variation in female body size was primarily explained by the first principal component (78.5% ± 1.53) and all three measurements (body length, femur length and pronotum width) had positive loadings. This principal component is referred to as 'female body size'. Female body size also differed significantly between *G. firmus*

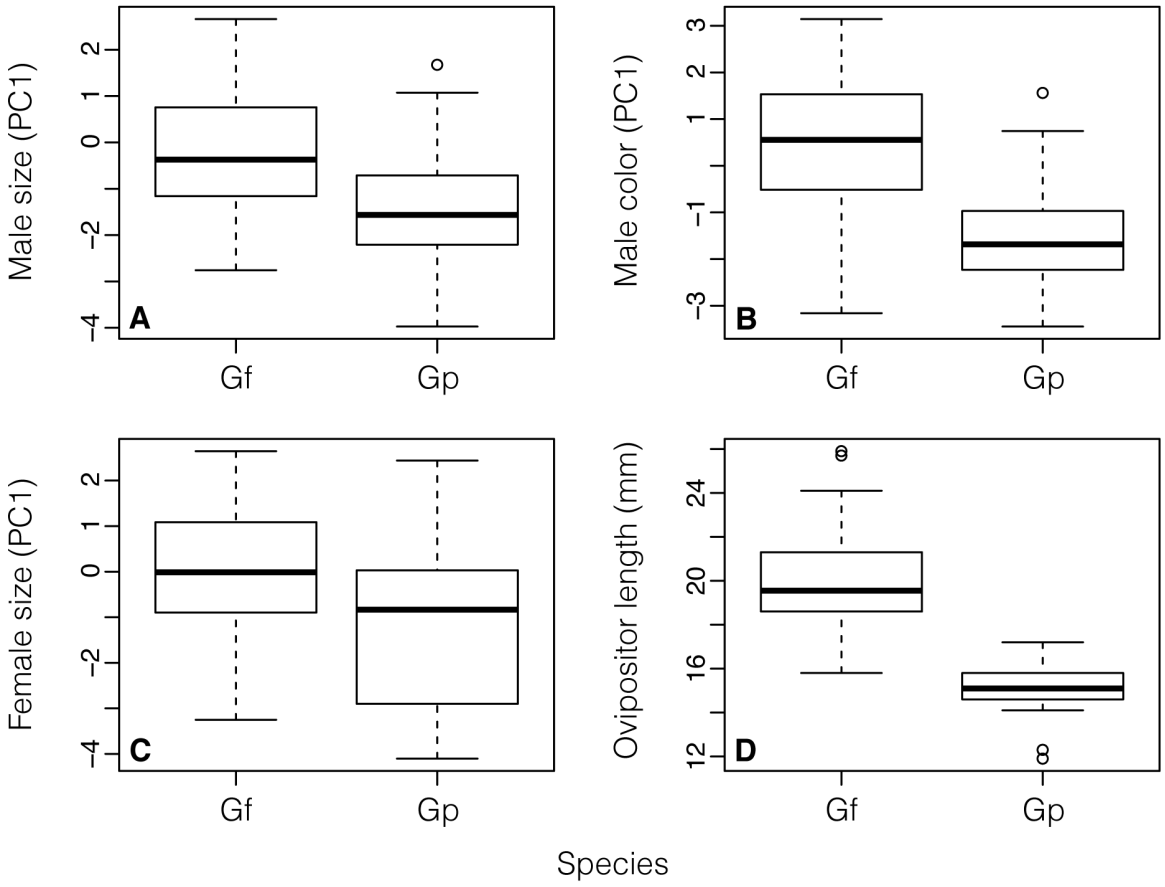


Figure 3.3 Morphological variation between allopatric populations of *Gryllus*. Boxplots of **A)** male body size, **B)** tegmina color, **C)** female body size and **D)** ovipositor length for allopatric populations of *G. firmus* (Gf: $N_{pops} = 5$, $N_{males} = 72$, $N_{females} = 56$) and *G. pennsylvanicus* (Gp: $N_{pops} = 4$, $N_{males} = 63$, $N_{females} = 21$).

(N = 56, -0.01 ± 1.38) and *G. pennsylvanicus* (N = 21, -1.10±1.85) (ANOVA: $F_{1,75} = 7.81$, P = 0.006) with *G. firmus* females having larger body sizes (Figure 3.3C). Ovipositor length showed the largest difference between *G. firmus* (N = 56, 20.08 ± 2.11) and *G. pennsylvanicus* (N = 21, 15.15 ± 1.39) (ANOVA: $F_{1,75} = 98.61$, P < 0.0001) with *G. firmus* having longer ovipositors (Figure 3.3D).

Together, male body size and tegmina color correctly classified most individuals from allopatric populations as either *G. firmus* or *G. pennsylvanicus* (91.1% correctly classified). Out of 72 *G. firmus* crickets, 11 were misclassified as *G. pennsylvanicus* and out of 63 *G. pennsylvanicus*, only 1 was misclassified as *G. firmus*. Pronotum width was the most important morphological character for classifying males, followed by tegmina hue (coefficients of linear discriminants: pronotum width = 0.979, tegmina hue = 0.795, body length = -0.387, femur length = -0.254, tegmina chroma = -0.122, tegmina brightness = 0.005). Female body size and ovipositor length together correctly classified all but a single *G. pennsylvanicus* individual as either *G. firmus* or *G. pennsylvanicus* (98.7% correctly classified). Ovipositor length was the most important morphological character for classifying females, followed by body length (coefficients of linear discriminants: ovipositor length = -0.600, pronotum width = 0.160, body length = 0.138, and femur length = -0.009).

Within the Pennsylvania hybrid zone, the majority of crickets were classified as either *G. firmus* or *G. pennsylvanicus* based on fuzzy c-means clustering (membership coefficients ≥ 0.90) (Figure 3.4A). Values of *r* ranging from 1.25 – 1.75 and 1.50 – 2.0 for males and females (respectively) classified approximately 15-50% of crickets as intermediate between the two parental types (Table 3.2). The distribution of cluster membership coefficients was bimodal for both males and females, with slightly more males classified as intermediate and an overall

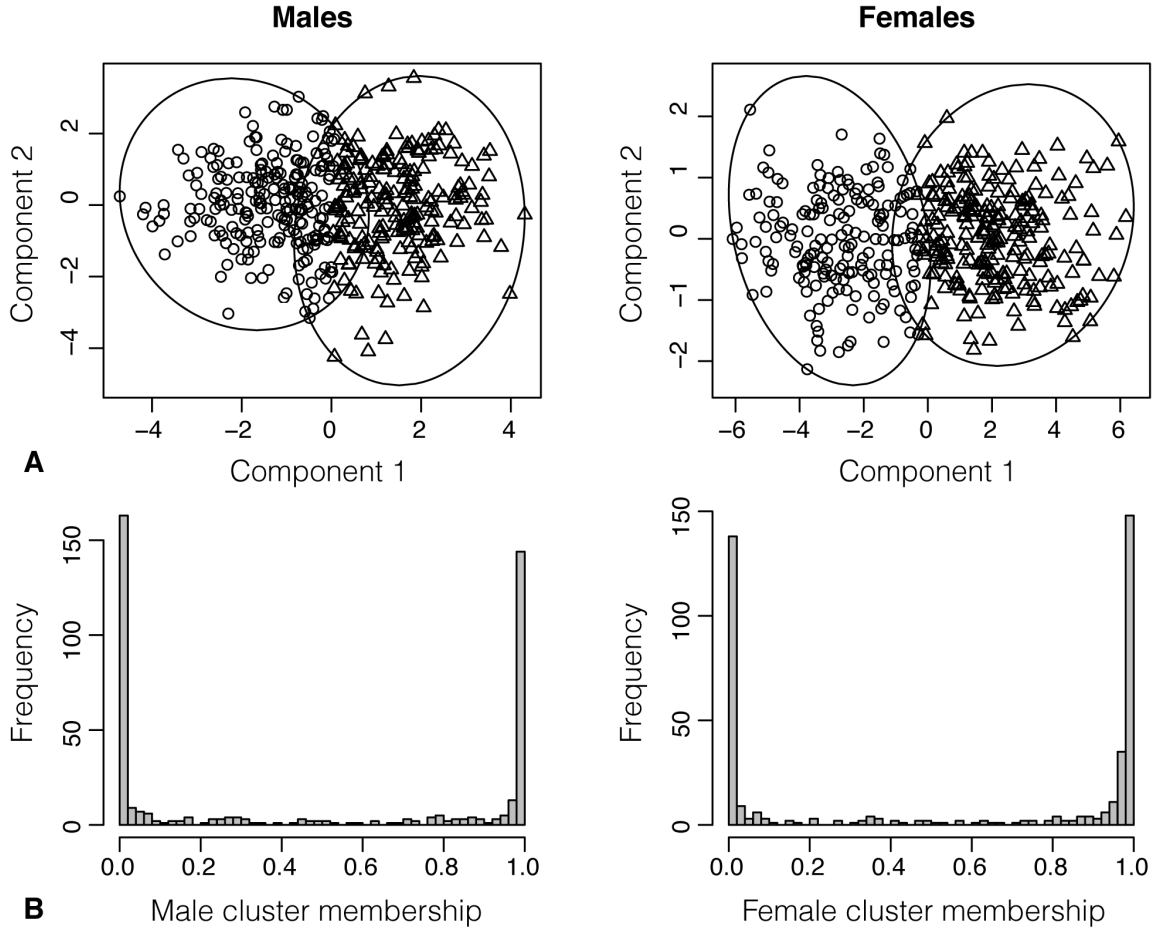


Figure 3.4 Morphological clusters for crickets from the Pennsylvania hybrid zone. A) Bivariate plot of morphological clusters ($k = 2$). Dissimilarities in morphological traits between individuals (males: body size and tegmina color; females: body size and ovipositor length) are calculated using squared Euclidean distances (fuzzy c-means) and represented by points in the plot using principal components. Ellipses indicate clusters of *G. pennsylvanicus* (open circles) and *G. firmus* (open triangles). **B)** Distribution of fuzzy cluster membership coefficients for males and females (*G. pennsylvanicus* ≤ 0.10 and *G. firmus* ≥ 0.90).

greater number of crickets classified as *G. pennsylvanicus* (Figure 3.4B) (see Figure S3.2 for the distribution at each sampling locality). The proportion of individuals in each sampling locality classified as *G. pennsylvanicus* (membership coefficient ≤ 0.10), *G. firmus* (membership coefficient ≥ 0.90) or intermediate (membership coefficient > 0.10 and < 0.90) is depicted in Figure 3.5. We found 2 collecting localities that were pure *G. pennsylvanicus*, 17 that were pure *G. firmus*. There was evidence of admixture (morphologically intermediate individuals or shared mtDNA haplotypes) in 32 localities that could overall be characterized as predominantly *G. firmus* (14) or *G. pennsylvanicus* (18). There were 36 localities that contained both parental types and admixed individuals (Table 3.1).

In 2010 we found 61 long-wing crickets at 20 localities. These crickets were predominantly *G. firmus* (based on morphological clustering) (Figure 3.6). Typically we collected only one or two long-winged crickets at a single locality (5-15%), but in several localities nearly a third had long-wings (CH, BV, BR) and at one locality, BG, nearly all the crickets we collected had long wings (77%).

Environmental predictors

Simple linear regressions- We found that natural vegetation ($F_{[1.77]} = 26.795$; $\beta = -0.037$; $R^2 = 0.258$; $P < 0.0001$), latitude ($F_{[1,77]} = 16.961$; $\beta = -3.254$; $R^2 = 0.180$; $P < 0.0001$), vegetation density ($F_{[1, 77]} = 13.054$; $\beta = -0.020$; $R^2 = 0.145$; $P = 0.0005$), annual temperature ($F_{[1,77]} = 11.570$; $\beta = 0.128$; $R^2 = 0.130$; $P = 0.0011$), and annual rainfall ($F_{[1, 77]} = 6.915$; $\beta = -0.013$; $R^2 = 0.082$; $P = 0.0103$) best predicted ovipositor length in the Pennsylvania hybrid zone (Figure 3.7). Percent sand was only a marginally significant predictor of ovipositor length ($F_{[1, 77]} = 4.110$; $R^2 = 0.050$; $P = 0.046$). Likewise, the same environmental variables were the best

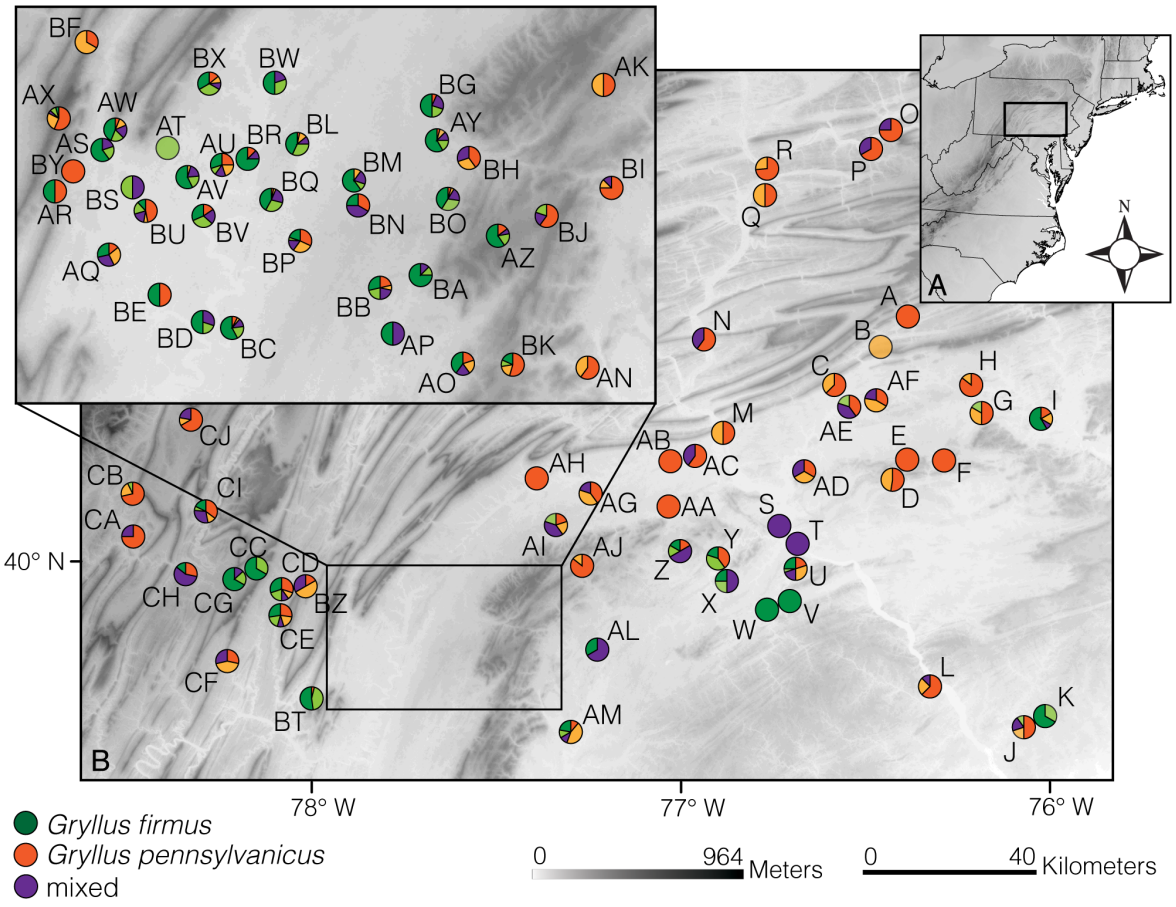


Figure 3.5 Distribution of *G. pennsylvanicus* and *G. firmus* in the Pennsylvania hybrid zone. Proportion of individuals within each collecting locality identified as either *G. pennsylvanicus* (orange) or *G. firmus* (green) based on morphological clustering. Species assignments are based on morphological clusters; individuals with membership coefficients of 0 or 1.0 are dark orange and dark green and individuals with membership coefficients ≤ 0.10 and > 0 or ≥ 0.90 and <1.0 are light orange and light green. Individuals with intermediate cluster coefficients < 0.90 and > 0.10 are purple.

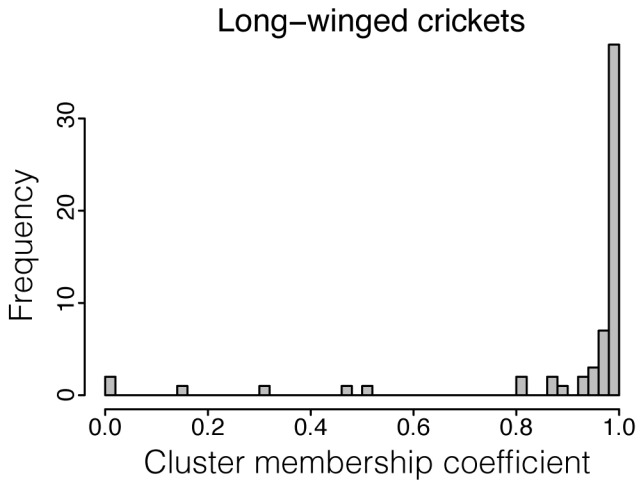


Figure 3.6 Distribution of cluster membership coefficients for long-wing morph crickets. There were a total of 61 long-wing morph crickets collected from 20 localities. The majority were *G. firmus*, based on morphological clustering. Localities where long-wing crickets were found are in bold in Table 3.1.

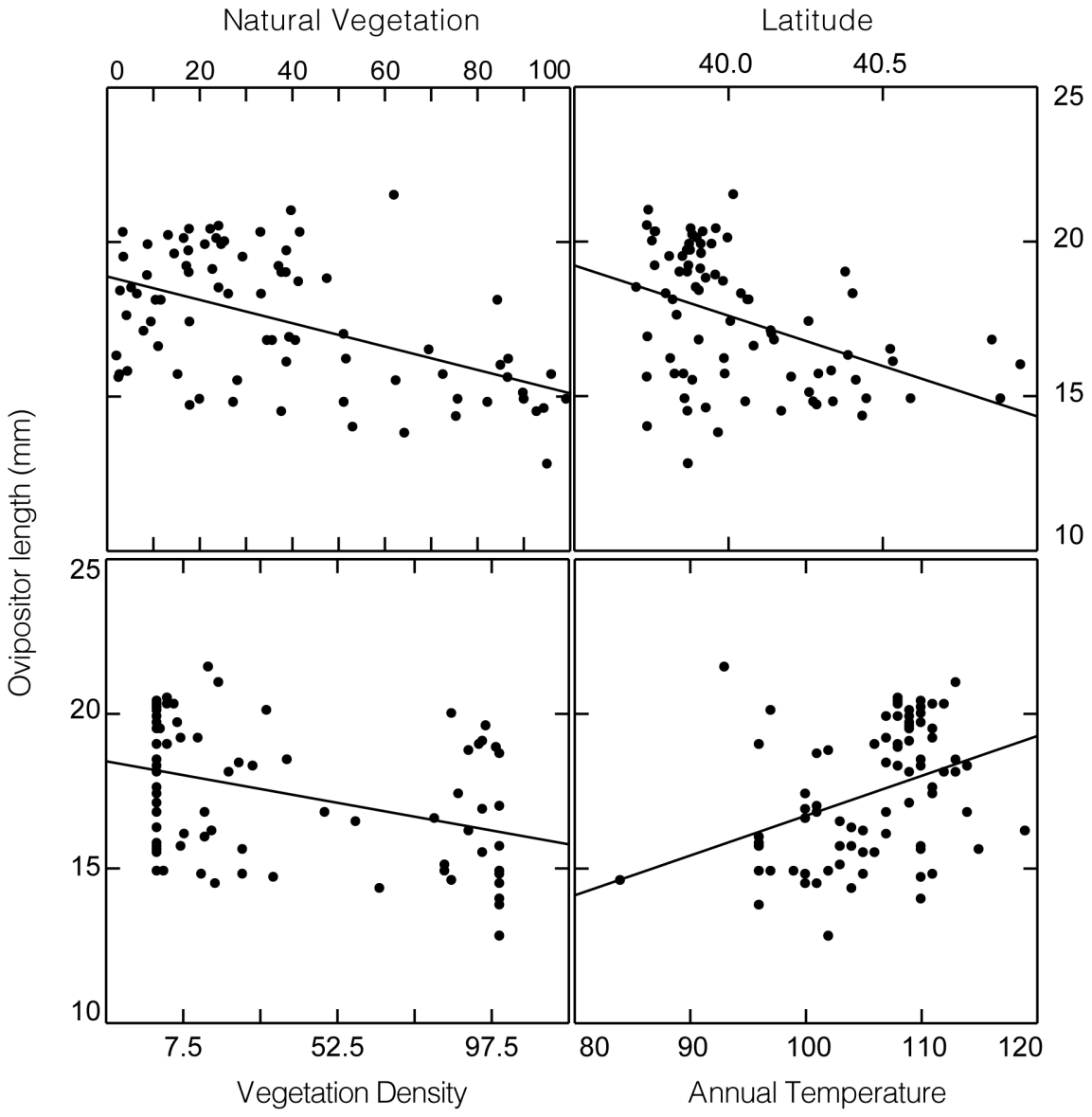


Figure 3.7 Simple linear regression of environmental variables. Scatterplots depicting the relationship between six environmental predictors and ovipositor length or clustering index.

predictors of the morphological clustering membership coefficients: latitude ($F_{[1,86]} = 31.165$; $\beta = -0.625$; $R^2 = 0.266$; $P < 0.0001$), natural vegetation ($F_{[1,86]} = 18.912$; $\beta = -0.005$; $R^2 = 0.180$; $P < 0.0001$), annual temperature ($F_{[1,86]} = 9.473$; $\beta = 0.018$; $R^2 = 0.099$; $P = 0.0028$), annual rainfall ($F_{[1,86]} = 5.998$; $\beta = -0.002$; $R^2 = 0.065$; $P = 0.016$) and vegetation density ($F_{[1,86]} = 5.004$; $\beta = -0.002$; $R^2 = 0.05$; $P5 = 0.028$). Percent sand and silt had only a marginally significant effect on morphological cluster (sand: $F_{[1,86]} = 4.529$; $R^2 = 0.050$; $P = 0.036$, silt: $F_{[1,86]} = 4.335$; $R^2 = 0.048$; $P = 0.040$). The following environmental variables did not significantly influence either ovipositor length or cluster membership when considered in simple linear regressions: elevation, human footprint, topographic complexity, and soil physical characteristics for ovipositor (silt, clay and organic content) and morphological cluster (clay, organic content) (Table S3.5).

Model selection: all possible models- Looking simultaneously at all environmental factors explaining ovipositor length and cluster membership, our model selection identified natural vegetation and latitude as key variables explaining cricket morphology (Table S3.6). The best model explaining ovipositor length included latitude, natural vegetation, and vegetation density as positive predictors, and also the interactions of natural vegetation with both latitude and vegetation density (Table 3.3). The best model explaining the cluster membership included latitude, natural vegetation, and organic matter as positive predictors, and also the interactions of natural vegetation with both latitude and organic matter (Table 3.3).

Discussion

The field cricket hybrid zone in Pennsylvania is a mosaic of genetically and morphologically distinct populations. The distribution of both genetic and morphological types is

Table 3.3 General linear model testing simultaneously the effects of environmental factors on ovipositor length and morphological clustering index for crickets from the Pennsylvania hybrid zone. Shorter ovipositors and lower morphological clustering membership coefficients represent *G. pennsylvanicus* crickets.

Term	Beta	Std Beta	Std Error	t-Ratio	VIF	*P*
Ovipositor Length						
Intercept	*186.183*	*0.000*	*26.779*	*6.95*	.	*<.0001*
Latitude	-4.181	-0.546	0.670	-6.24	1.277	<.0001
Veg. density	-0.013	-0.249	0.005	-2.66	1.466	0.010
Natural veg.	-0.016	-0.213	0.007	-2.16	1.622	0.034
Latitude*Natural veg.	0.050	0.215	0.021	2.45	1.287	0.017
Veg. density*Natural Veg.	-0.001	-0.272	0.000	-3.11	1.282	0.003
Morphological cluster						
Intercept	*29.478*	*0.000*	4.171	7.07	.	*<.0001*
Latitude	-0.717	-0.592	0.104	-6.89	1.276	<.0001
Organic Matter	-0.042	-0.180	0.018	-2.32	1.046	0.022
Natural Veg.	-0.004	-0.363	0.001	-4.59	1.079	<.0001
Latitude*Natural Veg.	0.012	0.331	0.003	3.900	1.244	0.0002
Organic Matter*Natural Veg.	0.001	0.238	0.001	3.000	1.089	0.003

Whole model statistics: ovipositor size ($F_{[5,73]} = 18.861$; $R^2 = 0.563$; $P < 0.0001$); morphological cluster. ($F_{[5,82]} = 18.084$; $R^2 = 0.524$; $P < 0.0001$).

highly heterogeneous and cannot be explained as a function of distance across the hybrid zone; there is no clear clinal pattern of variation for any trait, but rather a patchwork of populations. This is similar to the patterns seen in other regions of the hybrid zone in Virginia (Harrison and Arnold 1982) and Connecticut (Harrison 1986; Harrison and Rand 1989; Rand and Harrison 1989). Mosaic hybrid zones occur when the ecological settings and/or geography in the area of overlap are heterogeneous or complex, and species distributions are determined by environmental selection (e.g. Harrison 1986; Howard 1986; Harrison and Rand 1989; Bridle et al. 2001; Ross and Harrison 2002; Bierne et al. 2003; Vines et al. 2003; Ross et al. 2008). In Pennsylvania, we find an association between species distribution and natural habitat; *G. pennsylvanicus* occupies natural habitat along forest edges and natural clearings, while *G. firmus* occupies more disturbed areas in agricultural and suburban environments. Hybridization and introgression occur across patch boundaries; there is evidence of substantial admixture both in morphological characters and mtDNA, over a broad geographic area.

Broad scale distribution of G. firmus and G. pennsylvanicus: morphology and mtDNA

Our sampling of *G. firmus* and *G. pennsylvanicus* revealed the same four major mtDNA haplotype groups that were found by Willet et al. (2007) and Maroja et al. (2009a). *Gryllus pennsylvanicus* consists of two major clades (northern and southern) and *G. firmus* has a distinct southern clade, and a northern group that is distinguishable from the other three well-supported clades (Figure S3.1). The divide between northern and southern groups of each species appears to be centered within our study area in central Pennsylvania, where we find collecting localities containing all four mtDNA haplotypes (BB, AR, BK) (Figure 3.2C). Northern *G. pennsylvanicus* haplotypes can be found as far south as Durham, NC (DUR) and southern *G.*

pennsylvanicus haplotypes as far north as New Bloomfield, PA (NBL) (Figure 3.2B). A number of southern *G. firmus* haplotypes occur in southern Pennsylvania and New Jersey (Figure 3.2C) and there are a few northern *G. firmus* haplotypes in the Pennsylvania hybrid zone.

The relationship among the mtDNA groups has not been resolved with sequence data from the mtDNA COI gene. Willet et al. (1997) found five equally parsimonious tree topologies for these groups and in all cases either northern or southern *G. pennsylvanicus* clades were the basal group, with the southern *G. pennsylvanicus* clade having the greatest haplotype diversity. Additional genotyping of seven mtDNA SNPs identified two nucleotide positions in the ATPase6 and COIII genes that are shared among northern and southern *G. firmus* (Figure 3.2A). This suggests that an ancestral cricket lineage split into two daughter lineages, one that became either northern or southern *G. pennsylvanicus* and a second that split into the other *G. pennsylvanicus* clade and *G. firmus*. *Gryllus firmus* has subsequently diverged into northern and southern groups.

We also find that the hybrid zone extends further north than previously described (Harrison and Arnold 1982; Harrison 1986; Maroja et al. 2009a). In Rhode Island (PTJ) crickets that are *G. firmus*-like in morphology have *G. pennsylvanicus* mtDNA haplotypes, indicating there has been introgression of mtDNA. In Massachusetts (EST) we identified a mixed population that appears to contain both parental types. The majority of the crickets at this site have *G. firmus* morphology, but several crickets have smaller body size, darker tegmina, and a shorter ovipositor characteristic of *G. pennsylvanicus*. We also find that nearly half the crickets had *G. pennsylvanicus* haplotypes. The collecting locality was a wide beach, with sandy soils and short beach grasses, typical of where we would find pure *G. firmus* populations. To the north of this site we found only pure *G. pennsylvanicus* populations (sites SHA, POW). Thus,

Massachusetts may represent the northern range limit of *G. firmus*.

Patches of natural habitat maintain a mosaic structure

Throughout their ranges, both *G. pennsylvanicus* and *G. firmus* can be found in disturbed habitats along roadsides, in fields and pastures and around human settlement. Yet in the Pennsylvania hybrid zone we see an association between species distribution and natural habitat; *G. pennsylvanicus* occupies natural habitat along forest edges and clearings, while *G. firmus* occupies disturbed habitat near human settlement and agriculture (Table 3.3, Figure 3.7). In Pennsylvania study area, there is more natural habitat in the Appalachian Mountains to the north, which can explain why we also see a correlation between the distribution of *G. pennsylvanicus* and higher latitudes, greater vegetation density, lower temperatures and more rainfall. However, we also find *G. pennsylvanicus* crickets in patches of natural habitat further south along the Blue Ridge Mountains (*e.g.* AJ, AK, AN, BI) and near rivers, lakes and parks in the large gap between the Blue Ridge Mountains and the Reading Prong of the Northern Highlands (*e.g.* C, D, H, L, Z, Y). Likewise, disturbed habitats are typically in the lowland areas to the south, but valleys in the Appalachian Mountains are often intensely farmed and many are moderately populated. There are also corridors of human disturbance through the mountains along major highways (CH, CG, CC, CD, AZ, BJ).

Given that both cricket species seem well adapted to disturbed areas, it is unlikely that either performance in or preference for disturbed habitat restricts the distribution of *G. pennsylvanicus*. It is more likely that *G. firmus* is either less well-suited for the habitat characteristic of *G. pennsylvanicus*' range or that *G. firmus* is particularly well-suited for disturbed habitat and is a better colonizer. Both species are capable of dispersing over long

distances, but crickets of both species typically have short hind-wings and are incapable of flight. Daily movements such as feeding, reproduction and predator avoidance are accomplished by walking. But in years of high population densities individuals can be found with fully developed flight muscles and long hind-wings (Alexander 1968). Flight dimorphism is common in insects, and is thought to evolve in species that are adapted to disturbed or patchy environments for long-distance dispersal (Roff 1990). Although the development of long-winged morphs can have environmental triggers (*e.g.* temperature, population density, resources) (Harrison 1980), it is also heritable and populations vary in the proportion of long-winged individuals (Harrison 1979). An average of 4% of *G. pennsylvanicus* individuals are long-winged (Alexander 1968; Harrison 1979), whereas the frequency for *G. firmus* has been reported to be as high as 10-30% in some southern populations (Veazey et al. 1976). Wing dimorphism is thought to be particularly prevalent in *G. firmus* because its natural habitat is often highly disturbed and often ephemeral (e.g. sand dunes, beach grass and under shoreline debris) and may necessitate frequent dispersal among habitat patches. Indeed, of the 61 crickets we found with long hind-wings the majority (50 crickets) were *G. firmus* and only 2 were *G. pennsylvanicus* (the remainder were classified as hybrids based on morphological membership coefficients) (Figure 3.6). Wing dimorphism averaged 5-10% in *G. firmus* populations, but ranged as high as 30-75% at some collecting localities. In all cases, long-winged crickets were found at localities with high population densities and in disturbed habitats (Table 3.1). This suggests that *G. firmus* is both capable of thriving in disturbed habitats and may have a greater propensity for long-distance dispersal.

Overall, this is consistent with what appears to be a pattern of *G. firmus* expanding north into the Appalachian valleys of Pennsylvania. Harrison and Arnold (1982) identified mixed populations to the south in the Shenandoah Valley (a region of the Great Appalachian Valley that

extends through our study area) and Maroja et al. (2009a) reported three populations (COV, FRN, MOO) in the Appalachian Mountains west of the Shenandoah Valley where some individuals had *G. firmus* mtDNA haplotypes and *G. firmus*-like morphological traits (Figure 3.2B). We found a high proportion of crickets with both *G. firmus* morphology and mtDNA haplotypes in the southwestern corner of our study area. The intervening valleys are mostly farmland with moderate development. *Gryllus firmus* likely expanded north through these corridors and across the steep mountain ridges along roadways through natural water and wind gaps (CG, CH, CD, CE, BT). In some areas it appears that human disturbance has facilitated the persistence of *G. firmus* in otherwise heavily forested, natural habitats (CC and AZ).

Ecological barriers vary throughout the hybrid zone

The environmental variables maintaining hybrid zone structure are very different between Pennsylvania and the other regions of the hybrid zone. In Connecticut, we see a very clear association between species distributions and soil type. Ovipositor length can vary with soil type, allowing females to adjust how deeply eggs can be placed in the soil (Masaki 1979). Egg placement is a balance between protecting eggs from desiccation, cold temperatures and predation at lowers depths and the ability of the hatching nymph to emerge from the soil matrix (Bradford et al. 1993). Connecticut soils can vary over very short distances from loam to sand, and we see a pattern consistent with this hypothesis; *G. firmus* (longer ovipositor) occurs on sandy soils and *G. pennsylvanicus* on loam (Harrison 1986; Ross and Harrison 2002). In contrast, Pennsylvania soils are predominantly clay (≥ 20 % clay) and we saw no correlation between soil properties and species distributions.

In Virginia there is also no association between species distribution and soil type.

Instead, elevation and temperature appear to contribute to hybrid zone structure. *Gryllus firmus* from Virginia develop more slowly than *G. pennsylvanicus* (both in the field and in the lab) resulting in offset adult emergence (Harrison 1985). Genetic variation in life cycles (i.e. development time and voltinism) is common in species that extend over large climatic ranges and across elevation gradients (reviewed in Masaki 1983). There are likely climatic life cycle shifts in *G. firmus*; southern crickets develop quickly and have multiple generations per year, but in mid-latitudes development may slow to accommodate only one generation per year and at even higher latitudes shorter growing seasons may again favor faster development rate. This is consistent with *G. firmus* having a multivoltine lifecycle in the south, a univoltine life cycle with long development time in Virginia, and shorter univoltine development time in Connecticut (where this species emerges synchronously with *G. pennsylvanicus*) (Fulton 1952; Alexander 1968; Walker 1980; Harrison 1985). We collected crickets in Pennsylvania over two summers throughout August and September, and each summer we found both species with relatively equal frequency. No systematic survey of adult emergence across populations has been conducted, so we cannot rule out the possibility that there are environmental and/or genetic differences in development time. However, temporal isolation is less likely to be a major factor structuring the Pennsylvania hybrid zone.

Patterns of admixture suggest strong prezygotic barriers

In mosaic hybrid zones, the patchy distribution of parental types results in extensive contact throughout the zone. Hybridization and introgression occur across patch boundaries or in intermediate habitats. In the Pennsylvania hybrid zone, we see a patchy distribution of natural and disturbed habitat and a corresponding distribution of *G. pennsylvanicus* and *G. firmus*.

There are numerous opportunities for contact in areas where we see transitions in patch type: along mountains slopes, intersecting roadways and near encroaching human development. Despite these opportunities for hybridization, we find that the majority of collecting sites are predominantly composed of a single parental type and a few individuals with intermediate morphologies that may be admixed (most likely backcrosses) (Figure 3.5). Indeed, many of the crickets from sites with intermediate *G. firmus* morphologies had *G. pennsylvanicus* mtDNA haplotypes, suggesting that morphology is a good indicator of admixture. Each of these individual populations has an L-shaped distribution of morphological cluster membership (Figure S3.2), but the combination of these predominantly *G. firmus* and *G. pennsylvanicus* populations results in an overall bimodal distribution within the hybrid zone (Figure 3.4B). *Gryllus firmus* localities were more common in the Great Appalachian Valley between the Appalachians and the Blue Ridge, while *G. pennsylvanicus* were located mostly in the northeastern corner of our study area and in the large gap between the Reading Prong of the Northern Highlands and the Blue Ridge Mountains. A few collecting localities contained both parental types, and a number of sites appeared to be mixed (containing both parental types and morphologically intermediate individuals).

The topographic complexity of the region may also explain why the hybrid zone appears broader across the central Appalachian Mountains than early surveys of the hybrid zone suggested (Harrison and Arnold 1982; Willett et al. 1997; Maroja et al. 2009a). The sharp transitions between forested mountains and populated valleys increase the patchiness of natural habitat and could increase the extent of hybridization. In addition, increased human disturbance as a result of suburban expansion, agriculture and resource extraction is likely expanding the area of contact by increasing suitable habitat for *G. firmus*. Contact in some of these areas may even

be very recent. For instance, the occurrence of *G. firmus* along the Pennsylvania turnpike (CG, CH, CD) in relatively discrete locations suggests that *G. firmus* may have only begun occupying these high elevation sites in recent decades.

Although we find evidence of substantial admixture both in morphological characters and mtDNA over a broad geographic area, the two species remain distinct. Most individuals are morphologically like one or the other parental type (Figure 3.4B), and there are few intermediate individuals. Given that F1 hybrids are viable and fertile in the lab, this suggests that strong prezygotic barriers are operating in this portion of the hybrid zone (Jiggins and Mallet 2000), a pattern consistent with characterizations of other regions of the hybrid zone in Virginia (Harrison and Arnold 1982) and Connecticut (Harrison 1986; Harrison and Bogdanowicz 1997). Multiple prezygotic barriers have been identified between these species. Some of these barriers appear to be consistent throughout the hybrid zone, such as post-mating prezygotic barriers (Harrison 1983) and behavioral isolation (Harrison 1986; Harrison and Rand 1989; Maroja et al. 2009b). In contrast, the ecological barriers that may be maintaining the hybrid zone's mosaic structure appear to vary between geographic regions. Here, we find that the extent of natural habitat best explains the distribution of the two cricket species, whereas in Connecticut crickets are associated with different soil types and in Virginia crickets occur at different elevations. This variation can have important consequences for patterns of introgression within different regions of the hybrid zone. Species boundaries have been described as semipermeable, with permeability varying across different genomic regions. The permeability of species boundaries may also vary among different geographic areas and ecological contexts (Rand and Harrison 1989). Characterizing multiple regions within a hybrid zone is therefore critical for understanding hybrid zone dynamics, and gaining insights into the nature of species boundaries.

Acknowledgements

We would like to thank JD Larson for assistance with cricket collections, LS Maroja for providing DNA samples for the Pownal, VT population and the members of the Harrison Lab for thoughtful comments on this manuscript. This work was funded by a NSF grant to RGH and an American Association for University Women Fellowship to ELL.

References

Alexander, R. D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). Ann. Entomol. Soc. Am. 50:584-602.

Alexander, R. D. 1968. Life cycle origins speciation and related phenomena in crickets. Q. Rev. Biol. 43:1-41.

Andrés, J. A., E. L. Larson, S. M. Bogdanowicz, and R. G. Harrison. 2013. Assessing genomic differentiation in a pair of recently diverged field cricket species. Genetics in press.

Barton, N. H., and G. M. Hewitt. 1981. A chromosomal cline in the grasshopper *Podisma pedestris*. Evolution 35:1008-1018.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113-148.

Barton, N. H., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341:497-503.

Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard, F. Bonhomme, and P. David. 2003. Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. Mol. Ecol. 12:447-461.

Bradford, M. J., P. A. Guerette, and D. A. Roff. 1993. Testing hypotheses of adaptive variation in cricket ovipositor lengths. Oecologia 93:263-267.

Bridle, J. R., S. J. E. Baird, and R. K. Butlin. 2001. Spatial structure and habitat variation in a grasshopper hybrid zone. Evolution 55:1832-1843.

Doherty, J. A., and M. M. Storz. 1992. Calling song and selective phonotaxis in the field crickets, *Gryllus firmus* and *G. pennsylvanicus* (Orthoptera, Gryllidae). J. Insect Behav. 5:555-569.

Endler, J. A. 1977. Geographic Variation, Speciation and Clines. Princeton University Press, Princeton, NJ.

Endler, J. A. 1990. On the measurement and classification of colour in studies of animal colour patterns. Biol. J. Linn. Soc. 41:315-352.

Fulton, B. B. 1952. Speciation in the field cricket. Evolution 6:283-295.

Gompert, Z., L. K. Lucas, J. A. Fordyce, M. L. Forister, and C. C. Nice. 2010. Secondary contact between Lycaeides idas and L. melissa in the Rocky Mountains: extensive admixture and a patchy hybrid zone. Mol. Ecol. 19:3171-3192.

Harrison, R., and S. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36:535-552.

Harrison, R. G. 1979. Flight polymorphism in the field cricket *Gryllus pennsylvanicus*. Oecologia 40:125-132.

Harrison, R. G. 1980. Dispersal polymorphisms in insects. Annu. Rev. Ecol. Syst. 11:95-118.

Harrison, R. G. 1983. Barriers to gene exchange between closely related cricket species . I. laboratory hybridization studies. Evolution 37:245-251.

Harrison, R. G. 1985. Barriers to gene exchange between closely related cricket species . II. life-cycle variation and temporal isolation. Evolution 39:244-259.

Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56:337-349.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Pp. 69-128 *in* D. Futuyma, and J. Antonovics, eds. Oxford Surveys in Evolutionary Biology. Oxford University Press, New York, NY.

Harrison, R. G. 1993. Hybrids and hybrid zones: historical perspective. Pp. 3-12 *in* R. G. Harrison, ed. Hybrid Zones and the Evolutionary Process. Oxford University Press, Oxford.

Harrison, R. G., and S. M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51:493-505.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 111 - 133 *in* D. Otte, and J. A. Endler, eds. Speciation and its Consequences. Sinauer, Sunderland, MA.

Harrison, R. G., D. M. Rand, and W. C. Wheeler. 1987. Mitochondrial-DNA variation in field crickets across a narrow hybrid zone. Mol. Biol. Evol. 4:144-158.

Hewitt, G. M. 1988. Hybrid zones - natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158-167.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25:1965-1978.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogramm. Eng. Remote Sensing 73:337-341.

Howard, D. J. 1986. A zone of overlap and hybridization between two ground cricket species. Evolution 40:34-43.

Huaxing, L. 2008. Modeling terrain complexity. Pp. 159-176 *in* Q. Zhou, B. Lees, and G. Tang, eds. Advances in Digital Terrain Analysis. Springer, Berlin.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Jarvis, A., H. Reuter, A. Nelson, and E. Guevara. 2009. Hole-filled seamless SRM data V4.

Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends Ecol. Evol. 15:250-255.

Key, K. H. L. 1968. The concept of stasipatric speciation. Syst. Zool. 17:14-22.

Larson, E. L., G. L. Hume, J. A. Andrés, and R. G. Harrison. 2012. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. J. Evol. Biol. 25:174-186.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2002. cluster: Cluster analysis basics and extensions.

Maroja, L. S., J. A. Andrés, and R. G. Harrison. 2009a. Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. Evolution 63:2999-3015.

Maroja, L. S., J. A. Andrés, J. R. Walters, and R. G. Harrison. 2009b. Multiple barriers to gene exchange in a field cricket hybrid zone. Biol. J. Linn. Soc. 97:390-402.

Masaki, S. 1979. Climatic adaptation and species status in the lawn ground cricket. III. Ovipositor length. Oecologia 43:207-219.

Masaki, S. 1983. Climatic speciation in Japanese ground crickets. GeoJournal 7:483-490.

Montgomery, R. 2008. CLR. Queen's University, Kingston, ON.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol. Ecol. Res. 10:806-820.

Posada, D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253-1256.

Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone - mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432-449.

Roff, D. A. 1990. The Evolution of Flightlessness in Insects. Ecol. Monogr. 60:389-421.

Ross, C. L. 2000. Ecological and environmental factors affecting the structure and maintenance of a mosaic hybrid zone between the field crickets, *Gryllus firmus* and *G. pennsylvanicus*. Pp. 263. Ecology and Evolutionary Biology. Cornell University, Ithaca, NY.

Ross, C. L., J. H. Benedix Jr, C. Garcia, K. Lambeth, R. Perry, V. Selwyn, and D. J. Howard. 2008. Scale-independent criteria and scale-dependent agents determining the structure of a ground cricket mosaic hybrid zone (Allonemobius socius – Allonemobius fasciatus). Biol. J. Linn. Soc. 94:777-796.

Ross, C. L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56:2296-2312.

Ross, C. L., and R. G. Harrison. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115:53-68.

Sanderson, E. W., M. Jaiteh, M. A. Levy, K. H. Redford, A. V. Wannebo, and G. Woolmer. 2002. The human footprint and the last of the wild. Bioscience 52:891-904.

SAS. 2012. JMP, Version 10. SAS Institute Inc., Cary, NC.

Schaefer, J. A., and C. C. Wilson. 2002. The fuzzy structure of populations. Can. J. Zool. 80:2235-2241.

Searle, J. 1993. Chromosomal hybrid zones in Eutherian mammals. Pp. 309-353 *in* R. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, Oxford.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogentic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Ann. Entomol. Soc. Am. 87:651-701.

Staff, S. S. 2012. Soil Survey Geographic (SSURGO) Database [Pennsylvania State].

Szymura, J. M., and N. H. Barton. 1991. The genetic-structure of the hybrid zone between the fire-bellied toads *Bombina-bombina* and *B-variegata* - comparisons between transects and between loci. Evolution 45:237-261.

Team, R. D. C. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker. 2010. The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64:472-485.

USGS, and FAO. 2000. Global forest resources assessment (FRA 2000)- global forest canopy density image. Pp. http://edc2.usgs.gov/glcc/fao/forest_canopy_image.php.

Veazey, J. N., C. A. R. Kay, T. J. Walker, and W. H. W. Walker. 1976. Seasonal Abundance, Sex Ratio, and Macroptery of Field Crickets in Northern Florida. Ann. Entomol. Soc. Am. 69:374-380.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Springer, New York.

Vines, T. H., S. C. Kohler, A. Thiel, I. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton, and B. Nurnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B-variegata*. Evolution 57:1876-1888.

Walker, T. J. 1980. Mixed oviposition in individual females of *Gryllus firmus*: graded proportions of fast-developing and diapause eggs. Oecologia 47:291-298.

Willett, C. S., M. J. Ford, and R. G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79:484-494.

Wu, C. I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851-865.

Yanchukov, A., S. Hofman, J. M. Szymura, and S. V. Mezhzherin. 2006. Hybridization of *Bombina bombina* and *B-variegata* (Anura, Discoglossidae) at a sharp ecotone in western Ukraine: Comparisons across transects and over time. Evolution 60:583-600.

CHAPTER 4

DIFFERENTIAL INTROGRESSION IN A MOSAIC HYBRID ZONE BETWEEN THE FIELD CRICKETS *GRYLLUS FIRMUS* AND *G. PENNSLVANICUS*

Abstract

Hybrid zones are unique opportunities to functionally test the role of individual genomic regions in barriers to gene exchange between closely related species. Genomic regions with low gene flow between species reflect assortative mating, divergent selection or under-dominance. We use a large SNP dataset to estimate patterns of introgression in a hybrid zone between the field crickets *Gryllus pennsylvanicus* and *G. firmus,* which are isolated by multiple pre-zygotic barriers. We identified 36 genome regions that have significantly reduced gene flow within the hybrid zone. Several of these genes in these regions encode proteins that have functional roles consistent with barriers to fertilization (sperm/egg interactions) or habitat isolation (ovipositor length). We find no evidence that previously hypothesized candidate genes, those encoding seminal fluid proteins, have a significant role in reducing gene exchange within the cricket hybrid zone. Overall, we find evidence of historical introgression (extensive backcrossing), but few crickets that are products of immediate hybridization events (F1 hybrids). Strong prezygotic barriers to gene exchange maintain the hybrid zone, yet these barriers are incomplete. Identifying regions of the genome that are permeable to introgression is an important step towards understanding the genetic architecture of speciation and the evolutionary forces shaping divergence.

Introduction

The genomes of recently diverged species are mosaics; shared polymorphisms will characterize many regions, while other regions will have diverged in allele frequency as a result of random or selective lineage sorting (Harrison 1991; Wu 2001; Turner et al. 2005; Nosil et al. 2009). Some divergent regions may harbor genes contributing to intrinsic barriers between

species (or speciation phenotypes)(Shaw and Mullen 2011). Elucidating the genetic architecture of individual barriers (identifying "barrier genes") helps us to understand the evolutionary forces that lead to divergence and ultimately to speciation (Coyne and Orr 2004; Noor and Feder 2006; Nosil and Schluter 2011). However, we still know relatively little about the number of genes involved in barriers, how barrier genes are distributed throughout the genome, their interaction with other genes (gene products), and their consequences for gene flow. With the advent of increasingly efficient sequencing technologies we now have the capacity to look at many genes across the genome in organisms that reflect different modes of speciation and different stages of divergence (Noor and Feder 2006; Harrison 2010; Butlin et al. 2012; Nosil and Feder 2012).

The recent literature has focused primarily on the genetic architecture of barriers that evolve in the face of gene flow (Feder et al. 2012; Nosil and Feder 2012; Via 2012). It has been argued that the genetic architecture of barriers that develop in allopatry is less constrained and less informative, because the evolution of genetic divergence in allopatry is inevitable (Nosil and Feder 2012). However, the "inevitability" of allopatric speciation is what makes it an unquestionably important mode of diversification. The majority of speciation events likely involved allopatric divergence, or some combination of allopatry with periods or parapatry or sympatry (Coyne and Orr 2004). Given that the phases of speciation may vary with different modes of divergence (Nosil and Feder 2012), the genetic architecture of allopatric speciation warrants greater attention.

Allopatric divergence and subsequent secondary contact of recently diverged species provides a unique opportunity to dissect the genomic architecture of reproductive barriers. In hybrid zones, divergent lineages co-occur, and the extent to which genomic regions are exchanged can reveal where barrier genes reside (Barton and Hewitt 1985; Hewitt 1988;

Harrison 1990; Payseur 2010). Alleles that are equally fit in either genomic background should ultimately diffuse across the hybrid zone. In contrast, alleles at loci that are under divergent selection or selection against hybrid genotypes should exhibit limited introgression (Barton and Hewitt 1985; Harrison 1990; Rieseberg et al. 1999). We can estimate the strength of selection on a given locus by comparing the relative steepness of the change in allele or genotype frequencies across a gradient of hybridization (i.e., across a geographic transect or as a function of hybrid index) (Barton and Hewitt 1985; Szymura and Barton 1986; Gompert and Buerkle 2009). Neutral alleles will exhibit a sigmoidal cline, with cline width dependent on the rate of dispersal and the time since initial contact. Globally advantageous alleles will have a shallower cline, as they quickly spread across the hybrid zone. Alleles at barrier genes will have a characteristically steep cline (Key 1968; Bazykin 1969; Barton and Hewitt 1985). Thus, hybrid zones act as genomic sieves; neutral or advantageous regions will be freely exchanged between species whereas regions that contain barrier genes will not (Harrison 1990; Payseur 2010).

The hybrid zone literature has long documented patterns of differential introgression among genomic regions, and used these patterns to infer hybrid zone history, strength of selection, and number of genes maintaining species boundaries (reviewed in Harrison 1990, 1993; Rieseberg et al. 1999; Payseur 2010). More recently, there have been an increasing number of studies that have used differential introgression at multiple loci to identify regions with low gene flow between hybridizing taxa: sunflowers (Buerkle and Rieseberg 2001), poplar (Lexer et al. 2007; Lexer et al. 2010), spruce (Hamilton et al. 2012), buntings (Carling et al. 2008; Carling and Brumfield 2009), sculpin (Nolte et al. 2009), and house mouse (Teeter et al. 2008; Teeter et al. 2010; Dufková et al. 2011; Macholan et al. 2011; Janousek et al. 2012). With the advent of high-throughput sequencing, the genomes of closely related taxa can be scanned to

identify divergent regions (using either F_{ST} outliers or differences in allele frequencies) and to characterize patterns of introgression within and across hybrid zones. Divergent genome regions may be the result of divergent natural selection, and genome regions that introgress less than expected may be involved in reproductive isolation. However, ancestral polymorphism and lineage sorting also result in heterogeneous genomic divergence, particularly when recombination rates vary across the genome (Noor and Bennett 2009). Even genome regions under divergent selection may not necessarily be related to barriers between species (see Gompert et al. 2012a). Despite these caveats, the genome scan approach provides a catalogue of divergent gene regions that are candidates for harboring genes that contribute to reproductive barriers.

Here, we use a large single nucleotide polymorphism (SNP) dataset to estimate patterns of introgression across a hybrid zone between two field crickets, *G. pennsylvanicus* and *G. firmus,* that are isolated by multiple pre-zygotic barriers. We describe the distribution of genotypic classes across the hybrid zone and identify genomic regions with low gene flow, which may include loci responsible for reproductive barriers between these recently diverged species.

Field cricket Hybrid zone

The field crickets, *Gryllus pennsylvanicus* and *G. firmus* are morphologically very similar, but *G. firmus* is on average larger, with lighter tegmina and a relatively longer ovipositor. The two species, which diverged about 200,000 years ago (Willett et al. 1997; Maroja et al. 2009a), overlap in a hybrid zone that follows the eastern edge of the Appalachian mountains, extending from Massachusetts south into the Blue Ridge Mountains in Virginia (Harrison and Arnold 1982; Larson et al. 2013). Both species can be found in disturbed or grassy habitats, with *Gryllus*

firmus occupying lowland and coastal areas to the east and south, and *G. pennsylvanicus* in the uplands to the north and west (Harrison and Arnold 1982). Adult crickets appear in the late summer and early fall, females deposit eggs in the soil to overwinter, and nymphs emerge in spring (Alexander 1957, 1968).

Within the hybrid zone the distribution of parental types is a mosaic, with pure individuals of each species found in adjacent sites; both species, together with individuals of mixed ancestry, occur at some localities (Harrison 1986; Harrison et al. 1987; Harrison and Bogdanowicz 1997; Ross and Harrison 2002; Larson et al. 2013). Even where both parental species co-occur, few F1 hybrids are found (Harrison and Bogdanowicz 1997; Ross and Harrison 2002). The mosaic structure of the hybrid zone is a result of underlying habitat heterogeneity. In Connecticut, the patchy distribution of soil type (sand versus loam) dictates the structure of the zone (Harrison and Rand 1989; Rand and Harrison 1989; Ross and Harrison 2002). In Pennsylvania, species distributions are associated with the extent of natural vegetation, which is also correlated with elevation; the ridges of the Appalachian Mountains are characterized by more natural habitat occupied by *G. pennsylvanicus*, but human disturbance provides dispersal corridors for *G. firmus* from the lowlands (Larson et al. 2013).

Field inseminated *G. pennsylvanicus* females from mixed populations produce offspring sired primarily by conspecific males (Harrison 1986). Because there is no conspecific sperm precedence (Larson et al. 2012b), the observed pattern is presumably due to assortative mating. In laboratory no-choice mating trials, females of both species are reluctant to mate with males of the other species (Maroja et al. 2009b), suggesting that a combination of habitat isolation and behavioral cues underlie assortative mating. The cricket species exhibit a one-way incompatibility, in which crosses between *Gryllus firmus* females and *G. pennsylvanicus* males

never produce offspring (Harrison 1983). The incompatibility is a result of a combination of post-mating prezygotic barriers that reduce oviposition (Maroja et al. 2009b) and prevent stored *G. pennsylvanicus* sperm from properly fusing with *G. firmus* eggs (Larson et al. 2012b). In contrast, the reciprocal cross (*G. pennsylvanicus* female and *G. firmus* male) produces viable fertile offspring.

Despite differences in morphology, ecology and behavior, identifying genetic differences between the species has been difficult (allozymes, Harrison 1986; RFLPs and mtDNA, Harrison and Bogdanowicz 1997; Ross and Harrison 2002; nuclear gene introns, Broughton and Harrison 2003; seminal fluid proteins, Andrés et al. 2006; Andrés et al. 2008; Maroja et al. 2009a). A small number of RFLPs and a few seminal fluid proteins have been identified as divergent between allopatric populations (Harrison and Bogdanowicz 1997; Andrés et al. 2008; Maroja et al. 2009a). The overall similarity and on-going gene flow between the cricket species provides an opportunity to use patterns of differential introgression within the hybrid zone to identify genome regions that may underlie reproductive barriers. Recent transcriptome sequencing of the male accessory gland of each species identified > 6,000 SNPs, 10% of which had major allele frequency differences between the species (Andrés et al. 2013). Many of the genes expressed in the male accessory glands of *G. firmus* and *G. pennsylvanicus* are rapidly evolving under positive selection (Andrés et al. 2006; Braswell et al. 2006). Males transfer proteins secreted from the accessory gland to females as part of the ejaculate (packaged along with sperm in the spermatophore). Seminal fluid proteins (SFPs) are often rapidly evolving (Swanson and Vacquier 2002; Clark et al. 2006; Turner and Hoekstra 2008) and mediate critical steps in fertilization (Wolfner 2009). Divergence in SFPs could contribute to the one-way fertilization incompatibility between *G. firmus* females and *G. pennsylvanicus* males (Larson et al. 2012a;

Larson et al. 2012b). Through proteomic analyses of the male seminal fluid Andres et al. (2008) identified a list of unambiguous seminal fluid proteins that are transferred to females during copulation. Here, we use a subset of SNPs identified through comparisons of the male accessory gland transcriptome and SNPs from genes known to encode seminal fluid proteins to 1) quantify their differentiation between allopatric populations, 2) estimate genomic clines for a panel of the most highly differentiated SNPs, and 3) compare patterns of introgression between genes identified from the transcriptome scan and genes known to encode seminal fluid proteins.

Materials and Methods

Cricket Sampling

We used crickets collected from allopatric populations in the northern portion of the hybrid zone to validate putative SNPs identified from comparisons of transcriptome sequences (Andrés et al. 2013) and crickets from the hybrid zone in Pennsylvania to estimate introgression. For SNP validation we genotyped 71 crickets from three allopatric populations of each species (11-12 individuals per population) and nine crickets from a single mixed population that was outside of our focal study area (*G. pennsylvanicus*: ITH, SCR, SCO; *G. firmus*: GUI, TRI, PAR; mixed: MOO; described in Maroja et al. 2009 and Larson et al. 2013) (Table S4.1). To estimate introgression we genotyped 301 crickets from 36 localities within the hybrid zone in Pennsylvania described in Larson et al. (2013)(Table S4.2). Collection localities within the hybrid zone span the transition from the ridges and valleys of the Appalachian Mountains to the Great Appalachian Valley and coastal Piedmont region. We extracted genomic DNA from single adult femurs using the DNeasy Tissue Kit (QIAGEN Inc., Valencia, CA). The resulting DNA elution was diluted to 10 ng/μL.

SNP genotyping

We designed nine multiplexed assays targeting 232 SNPs in 181 contigs using the Sequenom MassARRAY platform. (Sequenom Inc., San Diego, CA). These assays were composed of 210 putative SNPs (168 contigs) identified from 454 and Illumina sequencing of two focal populations of each species (*G. pennsylvanicus*, Ithaca, NY; *G. firmus*, Guilford, CT) (Andrés et al. 2013) and 22 putative SNPs (15 contigs) from SFPs identified in *G. firmus* ESTs (Andrés et al. 2006; Andrés et al. 2008). We selected SNPs that were bi-allelic and had large frequency differences between our allopatric populations. The MassARRAY platform uses a multiplexed amplification of target DNA, followed by a single base extension of a primer immediately adjacent to the target SNP. The resulting product is analyzed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) and the mass difference of each possible SNP nucleotide allows unambiguous genotyping. Assays were designed using the MassARRAY Assay Design Software (Sequenom, San Diego, CA, USA) to multiplex up to 36 SNPs per well. For 34 contigs, assays were designed to encompass two or more SNPs in order to confirm that our genotyping assays had consistent genotype frequencies for each contig. For the remaining loci one SNP per contig was selected for genotyping. Amplicon primer sequences, expected product length, annealing temperature, extend primer sequence and target SNPs for each contig are listed in Table S4.3. Reactions were performed using iPLEX Gold chemistry at the Cornell Life Sciences Core Laboratories Center for Genomics. SNP genotypes were called using the Sequenom MassARRAY Typer v4.0 Analysis software and checked by eye. Assays that had poor amplification or peak resolution in our test panel of crickets (three allopatric populations of each species and one mixed population) were excluded from further analyses.

Admixture and Genomic Cline Analysis

We quantified admixture and estimated genomic clines using the R-package INTROGRESS (Gompert and Buerkle 2009; Gompert and Buerkle 2010). We first estimated the parental allele frequencies for each SNP with high quality amplification and genotype clustering in our allopatric populations (*G. pennsylvanicus*: ITH, SCR, SCO; *G. firmus*: GUI, TRI, PAR) using the function '*prepare.data*', and calculated the interspecific differentiation index (*D*) as:

$$D = \left| P_{Gf} - P_{Gp} \right|,$$

the absolute value of the allele frequency difference between the two species (see Renaut et al. 2010; Andrés et al. 2013) For our analyses we selected only SNPs for which $D \geq 0.80$ in comparisons of allopatric populations of the parental species; these represented 125 markers. We then quantified the ancestry of each cricket from the Pennsylvania hybrid zone by estimating a hybrid index, which is an average of the genome-wide admixture for a given individual, calculated as the proportion of alleles at all 125 markers that are inherited from *G. firmus*. Interspecific heterozygosity (the proportion of an individual's genome with alleles inherited from both parental populations) was estimated using the function '*calc.intsp.het*' and compared to the hybrid index to infer each individuals' hybrid class. Following Milne and Abbott (2008), crickets were defined as F1 individuals if they have an interspecific heterozygosity ≥ 85% and a hybrid index of 0.5, F2 or F3 individuals if they have an interspecific heterozygosity < 85% and a hybrid index between 0.25 and 0.75, or backcross individuals if they have an interspecific heterozygosity < 85% and a hybrid index ≤ 0.25 (backcross into *G. pennsylvanicus*) or ≥ 0.75 (backcross into *G. firmus*).

We constructed genomic clines using multinomial regression to predict, based on the hybrid index and interspecific heterozygosity, the probability of observing each of the three

possible genotypes (PP: homozygous *G. pennsylvanicus*, PF: heterozygous, and FF: homozygous *G. firmus*) at each marker. We compared the likelihoods of the regression model to a neutral model of introgression to identify markers that do not conform to expectations of neutral introgression. Our neutral model was constructed using 2,000 parametric simulations based on the observed genotype frequencies with the assumption that alleles at each locus are co-dominant (Gompert and Buerkle 2009). For all analyses, estimates of the hybrid index and interspecific heterozygosity are calculated taking into account allele frequencies in our parental populations, and significance thresholds were adjusted using the false discovery rate procedure (Benjamini and Hochberg 1995). The INTROGRESS output summarizes deviations from neutrality as either 1) excess or deficiency of homozygotes (*e.g.* PP+ PF+ FF-), which is consistent with direct selection, 2) excess of heterozygotes (*e.g.* PP- PF+ FF-), which is consistent with overdominance and 3) deficiency of heterozygotes (*e.g.* PP+ PF- FF+), which is consistent with underdominance, disruptive selection or assortative mating. Evidence for an excess or deficiency of homozygous and heterozygous genotypes at a given locus is based on the proportion of neutral simulations that yield a model with higher total probability of a given genotype than the model based on observed data (Gompert and Buerkle 2010). Deviation categories were assigned based on the INTROGRESS deviation category output, visual inspection of cline shape and observed genotype classes (see Macholan et al. 2011). Pairwise associations between different markers in our model output could indicate either physical linkage or epistasis. We tested for associations between each focal marker and other markers by adding the genotype of a predictor locus to our regression model and asking whether this information improved the fit of our model using Akaike information criterion (AIC). Smaller AIC values denote a higher likelihood and better model fitting. We calculated the difference in AIC values for our basic

model and our model with a predictor locus; large positive values indicate that the model with a predictor locus had improved fit over the basic model. The R code for tests of pairwise associations between alleles was supplied by Z. Gompert.

Results

Genotyping success

Of the initial 232 putative SNPs, 208 SNPs (89.7%) were successfully amplified in ≥ 95% of individuals from our test panel, and the majority of these (166 SNPs) had an amplification success ≥ 99% (Table 4.1). Visualization of individual PCR products revealed that most of the 24 SNPs that failed to amplify had introns that increased amplicon length, and likely performed poorly in assay multiplexes. Of the 208 SNPs successfully amplified, 23 SNPs (11.1%) had product sizes that were difficult to distinguish on the MALDI-TOF. Out of the 185 SNPs (155 contigs) that were both successfully amplified and had clear genotype clusters, 146 SNPs (125 contigs) had $D \geq 0.80$ between our six allopatric populations and 54 SNPs (46 contigs) of these had fixed differences between species ($D = 1.0$) (Table S4.4). We successfully genotyped > 1 SNP per contig for 15 contigs and all of these had similar genotype frequencies for different SNPs on the same contig. We restricted the remaining analyses to a single SNP per contig. Overall, there were 125 SNPs included in our analyses. For each contig, the total number of SNPs, the number of amino acid replacement SNPs per nonsynonymous site relative to the number of silent SNPs per synonymous site (pN/pS), sequence description, and functional annotation based on BLASTX are listed in Table S4.5 (Andrés et al. 2013). Overall, these SNPs had very high genotyping success for all populations; allopatric *G. firmus* 99.95% (4,500 possible genotypes); allopatric *G. pennsylvanicus* 99.95% (4,375 possible genotypes), our test

Table 4.1 Amplification and genotyping success of 232 SNP markers (181 contigs) for a panel of 71 crickets from six allopatric populations (*G. pennsylvanicus* = 35, *G. firmus* =36) and nine crickets from a single mixed population. *D* represents the interspecific differentiation index.

Category	**SNPs**	**%**	**Contigs**	**%**
Amplification				
≥ 95% of individuals	208	89.7	169	93.4
< 95% of individuals	24	10.3	12	6.6
Genotyping				
High quality genotype clusters	185	88.9	154	91.1
Poor genotype resolution	23	11.1	15	8.9
Genotyping results (185 SNPs)				
$D < 0.80$	39	21.1	29	18.8
$D \geq 0.80$	92	49.7	79	51.3
$D = 1.0$	54	29.2	46	29.9
		Total	**125**	**81.2**

mixed population 99.82% (1,125 possible genotypes) population and the hybrid zone 99.63% (37,625 possible genotypes).

Genomic structure

The majority (~93%) of individuals within the Pennsylvania hybrid zone had low interspecific heterozygosity (heterozygous for < 20% of markers) and the hybrid zone had a distinctly bimodal distribution of hybrid indices (Figure 4.1). Only one cricket could be considered an F1 hybrid (≥85% heterozygosity and a hybrid index of 0.50), eight crickets were identified as multigenerational hybrids (F2, F3, or F4 individuals), and the remainder was classified as either backcrosses into *G. pennsylvanicus* or *G. firmus*. Collections from the majority of localities were either predominantly *G. pennsylvanicus* or *G. firmus*, and all of these except AK also contained backcrossed individuals (Figure 4.2). There were eleven localities that contained both parental types (BZ, CE, AX, AR, BU, BK, CI, Y, J, AW, AZ), but in only seven localities were there multi-generation hybrids (CD, AX, BK, BU, BV, L, J). Ovipositor length, the morphological trait that best resolves the two species, shows a strong correlation with the hybrid index; crickets with a low hybrid index (*G. pennsylvanicus*) have shorter ovipositors and of the nine female crickets with intermediate hybrid indices, eight have intermediate sized ovipositors (one cricket body was damaged and her ovipositor was not measured) (Figure 4.3).

Genomic cline analyses

The extent of introgression varied greatly among the 125 markers; some markers showed very few heterozygotes, whereas for others, alleles of one species were found in crickets with overall hybrid indices characteristic of the other species (Figure 4.4). Of the 125 markers,

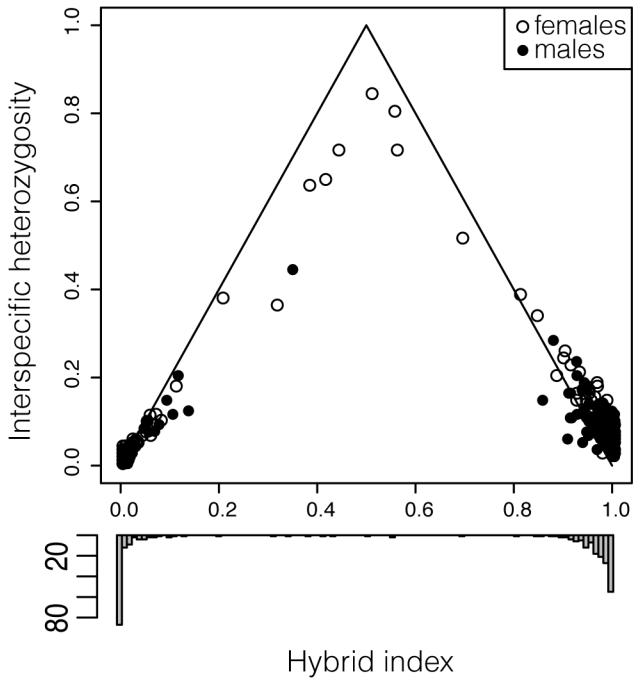


Figure 4.1 Interspecific heterozygosity plotted against hybrid index for females (open circles) and males (closed circles) from the Pennsylvania hybrid zone. Hybrid indices of zero and one represent individuals that are pure *G. pennsylvanicus* and *G. firmus*, respectively. F1 hybrids have an interspecific heterozygosity of ≥ 85% and a hybrid index of 0.5, multi-generation hybrids have an interspecific heterozygosity of < 85% and intermediate hybrid indices (> 25% and < 75%) and backcrosses into *G. pennsylvanicus* and *G. firmus* have an interspecific heterozygosity < 85% and hybrid indices < 25% or > 75% (see text for further details).

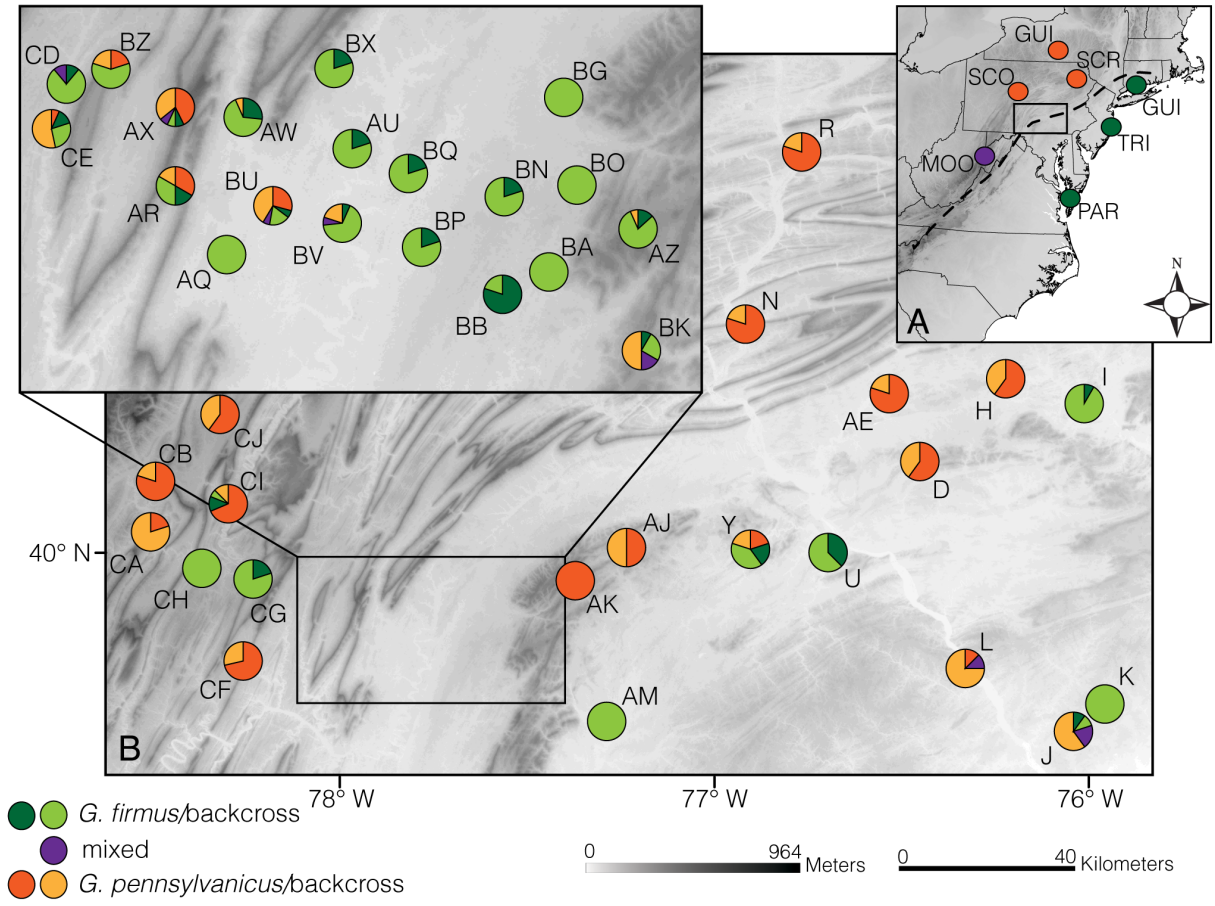


Figure 4.2 Distribution of hybrid classes for *G. pennsylvanicus* and *G. firmus* crickets in the Pennsylvania hybrid zone. Each pie diagram shows the proportion of parental and hybrid classes at a single sampling locality. Pure *G. pennsylvanicus* and *G. firmus* crickets are represented as dark orange and green, backcrosses into each parental type are lighter orange and green and multi-generation hybrids are purple. Letters refer to the detailed location information in Tables S1 and S2. (A) Allopatric sampling locations and the location of the hybrid zone (dashed line). *Gryllus firmus* occurs south and east of the hybrid zone and *G. pennsylvanicus* occurs north and west. The rectangle highlights the location of the study area. (B) Detailed view of the sampling localities within the hybrid zone.

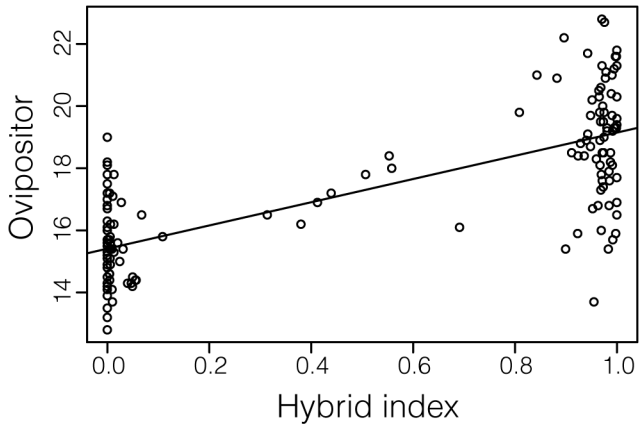


Figure 4.3 Relationship between the hybrid index and female ovipositor length. The ovipositor is the morphological trait that best distinguishes the two species; *G. firmus* on average has a longer ovipositor than *G. pennsylvanicus*.

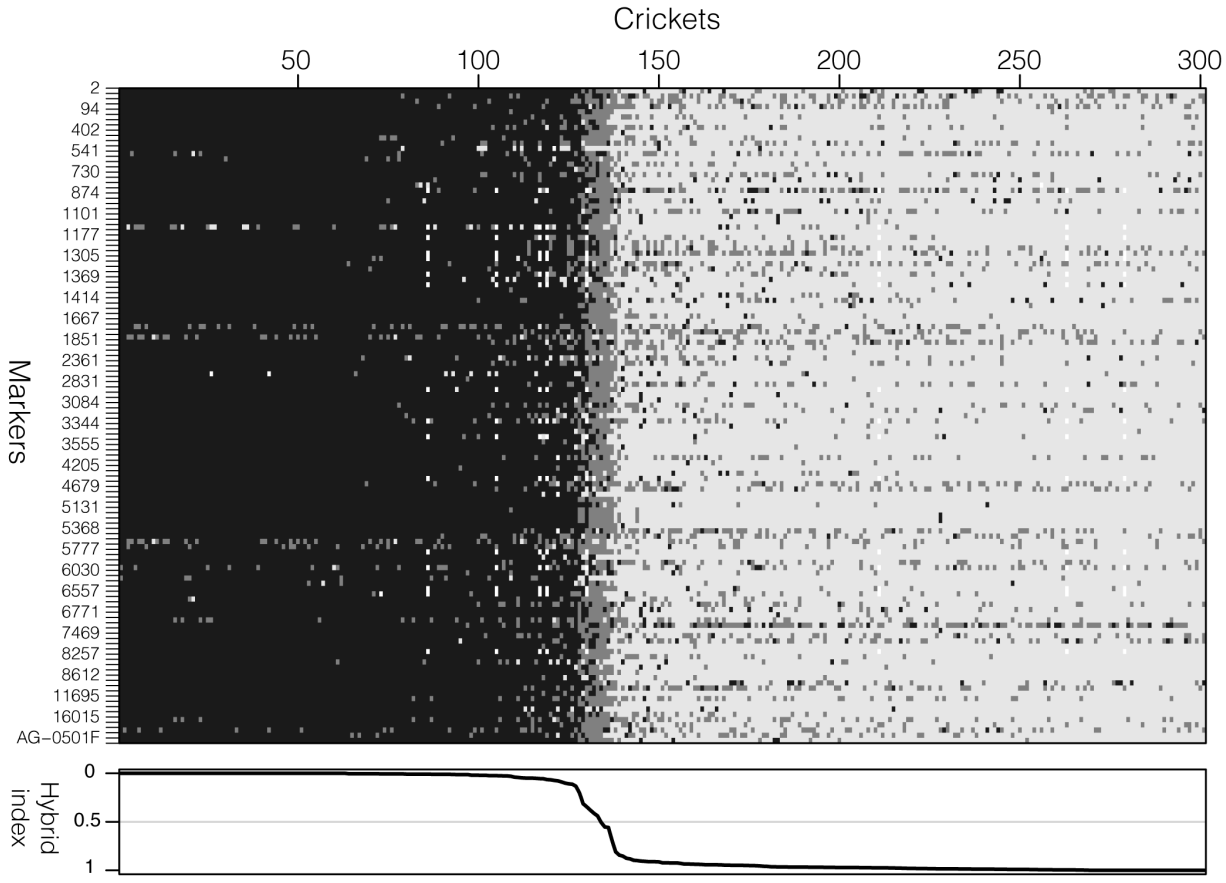


Figure 4.4 Overview of genotyping results for 125 SNP markers (ordered by contig number) in 301 individuals from the Pennsylvania hybrid zone. Each rectangle represents an individual's genotype at a given locus. Homozygous *G. pennsylvanicus* genotypes are dark grey, heterozygotes are medium grey, homozygous *G. firmus* genotypes are light grey, and white represents missing data. At the bottom is a visual representation of each individual's hybrid index, calculated as the proportion of the genome inherited from *G. firmus*.

42 (33.6 %) did not show patterns that deviated significantly from neutral expectations. We categorized the remaining 83 markers as deviating from neutral expectations based on the excess (+) or deficit (-) of homozygous and heterozygous genotypes and overall cline shape (Table 4.2). Thirty-five markers (28 %) had patterns of introgression consistent with directional selection. These were either loci that that had an excess of *G. pennsylvanicus* alleles (PP+ and FP+) in crickets with a *G. firmus* genomic background (25 markers) or that had an excess of *G. firmus* alleles (FF+ and FP+) in crickets with a *G. pennsylvanicus* background (10 markers). Twelve markers (9.6 %) had an excess of heterozygotes, consistent with overdominance or heterozygote advantage. Thirty-six markers (28.8%) had fewer heterozygotes than expected, and an excess of one or both homozygous genotype classes, consistent with underdominance, disruptive selection, or assortative mating. Ten of these markers (8 %) had a significant excess of both homozygous genotype classes. A summary of the functional annotation groups for these 36 markers is provided in Table 4.3. Figure 4.5 provides a summary of all 125 genomic clines and individual genomic clines for each marker are shown in Figure S4.1. A high proportion of markers showed evidence for two-locus interactions. For many markers, adding a second locus as a component in our multinomial regression to predict the genotype at our focal locus, greatly improved the model fit, which can be seen in the high ΔAIC values (dark grey) in Figure S4.2. This suggests that many of these markers may be associated, either through physical linkage or epistatic interactions.

Comparison with seminal fluid proteins

We successfully genotyped SNPs in 13 out of 15 genes known to encode seminal fluid proteins (SFPs) (Table 4.4). Four of these appear to be monomorphic, however it is not possible

Table 4.2 Results of genomic cline analysis for 125 markers genotyped in the Pennsylvania portion of the hybrid zone between *G. firmus* and *G. pennsylvanicus*. Contigs in bold represent markers that had significantly reduced introgression within the hybrid zone.

Contig	*D*[1]	sex-linked	lnL [2]	*P*[3]		Genotypes[4]			Dev.[5]
2	0.917	-	15.599	0	*	PP+	PF-	FF	P → F
80	0.847	no	5.506	0.023	*	PP	PF+	FF-	het +
87	0.875	no	3.913	0.095		PP-	PF	FF	n.s.
90	0.903	no	8.078	0.013	*	PP+	PF+	FF-	P → F
94	**0.917**	**-**	31.789	0	*	**PP**	**PF-**	**FF+**	**het -**
136	0.944	no	7.794	0.015	*	PP	PF	FF	P → F
211	**0.972**	**no**	10.093	0.003	*	**PP**	**PF-**	**FF+**	**het -**
367	0.958	no	2.198	0.335		PP	PF+	FF-	n.s.
402	**1**	**no**	6.696	0.027	*	**PP+**	**PF-**	**FF**	**het -**
425	0.986	no	5.958	0.053		PP-	PF	FF	n.s.
432	0.917	no	4.137	0.103		PP-	PF+	FF	n.s.
518	0.971	-	58.854	0	*	PP	PF-	FF+	F → P
541	0.902	no	8.945	0.004	*	PP	PF+	FF-	het +
580	0.957	-	5.573	0.061		PP+	PF-	FF+	n.s.
618	1	no	3.399	0.191		PP	PF+	FF-	n.s.
726	1	-	-0.556	0.847		PP	PF-	FF+	n.s.
730	0.958	no	16.204	0	*	PP	PF+	FF-	P → F
755	0.889	-	27.734	0	*	PP+	PF-	FF+	P → F
827	1	-	0.855	0.610		PP	PF-	FF+	n.s.
855	0.847	no	21.874	0	*	PP+	PF+	FF-	P → F
874	**1**	**-**	11.573	0.001	*	**PP**	**PF-**	**FF+**	**het -**
937	1	no	47.885	0	*	PP+	PF	FF-	P → F
963	**1**	**-**	12.761	0.001	*	**PP**	**PF-**	**FF+**	**het -**
1032	0.875	no	0.599	0.796		PP	PF	FF	n.s.
1101	**0.972**	**-**	9.035	0.010	*	**PP+**	**PF-**	**FF+**	**het -**
1121	**0.931**	**-**	17.067	0.001	*	**PP**	**PF-**	**FF+**	**het -**
1145	0.929	no	29.137	0	*	PP-	PF	FF+	F → P
1147	**1**	**-**	14.706	0.001	*	**PP**	**PF-**	**FF**	**het -**
1177	1	no	6.829	0.022	*	PP	PF+	FF-	P → F
1231	1	no	3.744	0.170		PP	PF+	FF-	n.s.
1234	0.889	no	1.929	0.356		PP	PF	FF	n.s.
1275	0.903	no	13.461	0.002	*	PP+	PF+	FF-	P → F
1305	**0.873**	**-**	19.857	0	*	**PP+**	**PF-**	**FF+**	**het -**
1309	0.944	no	13.200	0.001	*	PP	PF+	FF-	het +
1313	0.986	no	11.735	0.002	*	PP	PF+	FF-	P → F
1341	0.986	no	4.980	0.076		PP	PF+	FF-	n.s.
1369	1	-	60.382	0	*	PP	PF-	FF+	F → P
1372	**0.972**	**-**	15.619	0	*	**PP+**	**PF-**	**FF**	**het -**
1374	0.972	-	6.545	0.040		PP	PF-	FF+	n.s.
1412	1	no	8.510	0.010	*	PP+	PF	FF-	P → F

Table 4.2 (Continued)

1414	0.903	no	10.869	0.002	*	PP+	PF	FF	P → F
1513	0.972	no	3.780	0.144		PP+	PF-	FF	n.s.
1539	**1**	**no**	14.627	0	*	**PP**	**PF-**	**FF**	**het -**
1555	1	no	1.833	0.425		PP	PF+	FF-	n.s.
1667	1	-	3.302	0.210		PP+	PF-	FF	n.s.
1724	0.914	no	37.144	0	*	PP	PF+	FF-	het +
1774	0.986	no	62.035	0	*	PP+	PF+	FF-	P → F
1790	0.845	no	12.473	0.001	*	PP-	PF	FF+	F → P
1851	0.944	no	33.653	0	*	PP+	PF+	FF-	P → F
2100	1	no	8.987	0.010	*	PP	PF+	FF-	P → F
2182	1	-	5.510	0.060		PP+	PF	FF	n.s.
2271	0.917	no	21.919	0	*	PP-	PF	FF+	F → P
2361	0.944	no	0.777	0.682		PP	PF	FF-	n.s.
2467	**1**	**-**	10.992	0.001	*	**PP**	**PF-**	**FF+**	**het -**
2570	0.972	no	25.703	0	*	PP-	PF-	FF+	F → P
2733	1	no	3.303	0.198		PP	PF	FF	n.s.
2831	0.917	no	3.231	0.172		PP	PF	FF	n.s.
2833	1	-	-0.743	0.856		PP	PF-	FF	n.s.
2864	**0.972**	**-**	9.196	0.006	*	**PP+**	**PF-**	**FF+**	**het -**
2989	1	--	1.510	0.477		PP+	PF-	FF	n.s.
3084	0.931	no	1.634	0.393		PP-	PF	FF	n.s.
3136	0.847	-	13.361	0	*	PP	PF-	FF+	P → F
3182	1		12.515	0.001	*	PP-	PF	FF+	F → P
3268	0.972	no	5.909	0.046		PP-	PF+	FF-	n.s.
3344	1	no	5.275	0.075		PP+	PF-	FF	n.s.
3422	**0.972**	**-**	6.930	0.023	*	**PP**	**PF-**	**FF+**	**het -**
3432	**1**	**no**	10.098	0.001	*	**PP+**	**PF-**	**FF**	**het -**
3528	**0.972**	**no**	13.818	0.001	*	**PP**	**PF-**	**FF+**	**het -**
3555	1	-	-0.709	0.866		PP	PF-	FF+	n.s.
3838	**1**	**-**	12.822	0.001	*	**PP**	**PF-**	**FF+**	**het -**
3843	0.972	no	7.423	0.021	*	PP	PF+	FF-	het +
3968	1	-	3.172	0.211		PP+	PF	FF	n.s.
4205	**1**	**no**	10.045	0.002	*	**PP**	**PF**	**FF+**	**het -**
4361	0.986	-	10.634	0.002	*	PP+	PF	FF-	P → F
4450	1	no	1.109	0.531		PP+	PF-	FF	n.s.
4655	0.972	no	27.571	0	*	PP-	PF+	FF-	het +
4679	0.944	no	12.907	0.001	*	PP	PF+	FF-	het +
4913	**1**	**-**	22.488	0	*	**PP**	**PF-**	**FF+**	**het -**
5021	**1**	**-**	11.685	0.002	*	**PP**	**PF-**	**FF+**	**het -**
5052	1	-	5.149	0.077		PP+	PF-	FF	n.s.
5131	**1**	**-**	9.111	0.010	*	**PP**	**PF-**	**FF+**	**het -**

Table 4.2 (Continued)

5136	**0.972**	**-**	9.583	0.006	*	**PP**	**PF-**	**FF+**	**het -**
5177	**0.931**	**-**	23.735	0	*	**PP**	**PF-**	**FF+**	**het -**
5214	**1**	**no**	10.724	0.002	*	**PP+**	**PF-**	**FF+**	**het -**
5368	0.972	no	42.057	0	*	PP+	PF+	FF-	P → F
5556	0.917	no	3.536	0.148		PP-	PF+	FF	n.s.
5711	0.845	no	12.449	0.001	*	PP-	PF	FF+	F → P
5727	0.902	no	4.332	0.105		PP-	PF	FF	n.s.
5777	**1**	**-**	13.927	0	*	**PP**	**PF**	**FF**	**het -**
5961	**1**	**no**	29.641	0	*	**PP**	**PF-**	**FF+**	**het -**
6023	0.931	no	2.102	0.335		PP	PF+	FF-	n.s.
6026	0.916	no	10.898	0	*	PP	PF+	FF-	het +
6030	0.986	no	5.832	0.056		PP+	PF-	FF+	n.s.
6057	0.929	no	26.391	0	*	PP	PF-	FF+	F → P
6128	**0.902**	**no**	18.210	0	*	**PP+**	**PF-**	**FF+**	**het -**
6271	**0.931**	**no**	12.233	0.001	*	**PP**	**PF-**	**FF+**	**het -**
6557	0.986	no	-0.269	0.802		PP	PF	FF	n.s.
6571	0.958	no	3.700	0.175		PP	PF	FF	n.s.
6579	0.931	no	7.022	0.027	*	PP	PF	FF	het +
6718	1	-	31.397	0	*	PP+	PF	FF-	P → F
6771	**0.931**	**-**	16.984	0	*	**PP**	**PF-**	**FF+**	**het -**
7046	0.873	no	2.840	0.241		PP	PF	FF-	n.s.
7153	0.806	no	37.487	0	*	PP+	PF+	FF-	P → F
7164	**0.972**	**-**	9.590	0.006	*	**PP**	**PF-**	**FF+**	**het -**
7469	0.889	no	1.751	0.409		PP	PF	FF	n.s.
7566	0.875	no	6.226	0.030	*	PP+	PF	FF-	P → F
8026	1	no	5.959	0.051		PP	PF-	FF+	n.s.
8229	**1**	**-**	15.301	0	*	**PP**	**PF**	**FF+**	**het -**
8257	1	no	8.232	0.015	*	PP	PF+	FF-	P → F
8322	0.971	no	17.654	0	*	PP	PF-	FF+	F → P
8354	1	-	0.074	0.752		PP	PF-	FF	n.s.
8375	**1**	**no**	10.344	0.003	*	**PP**	**PF**	**FF**	**het -**
8612	**0.986**	**no**	11.025	0.004	*	**PP**	**PF-**	**FF+**	**het -**
9839	0.917	no	22.389	0	*	PP+	PF-	FF	P → F
9851	0.972	no	48.441	0	*	PP+	PF+	FF-	P → F
10368	**1**	**no**	8.536	0.009	*	**PP**	**PF-**	**FF+**	**het -**
11695	0.986	no	19.079	0	*	PP-	PF+	FF-	het +
12397	**1**	**-**	11.769	0.001	*	**PP**	**PF-**	**FF+**	**het -**
14713	**1**	-	46.515	0	*	**PP+**	**PF-**	**FF+**	**het -**
14937	1	no	3.589	0.168		PP	PF+	FF-	n.s.
16015	0.845	no	2.538	0.327		PP	PF	FF	n.s.
AG-0148P	0.875	no	13.225	0.001	*	PP-	PF-	FF+	P → F

Table 4.2 (Continued)

AG-0203P	0.888	no	9.152	0.006	*	PP-	PF+	FF	het +
AG-0383F	0.972	no	6.912	0.028	*	PP-	PF+	FF-	het +
AG-0501F	1	-	3.829	0.155		PP+	PF	FF	n.s.

[1] interspecific differentiation index; [2] likelihood ratio; [3] probability of departure from neutrality following false discovery rate correction (Benjamini and Hockber 1995); [4] over (+) or underrepresentation (-) of observed genotypes and [5] inferred deviation category.

Table 4.3 Summary of the functional annotation grouped into general protein classes for 36 markers that have reduced introgression.

Protein Class	Markers
Unknown function	14
Cytoskeleton related proteins	5
Protein binding/interactions	5
Growth	3
Translation initiation/regulation	2
GTPase related	2
Protease inhibitor	1
Other	4
Total	36

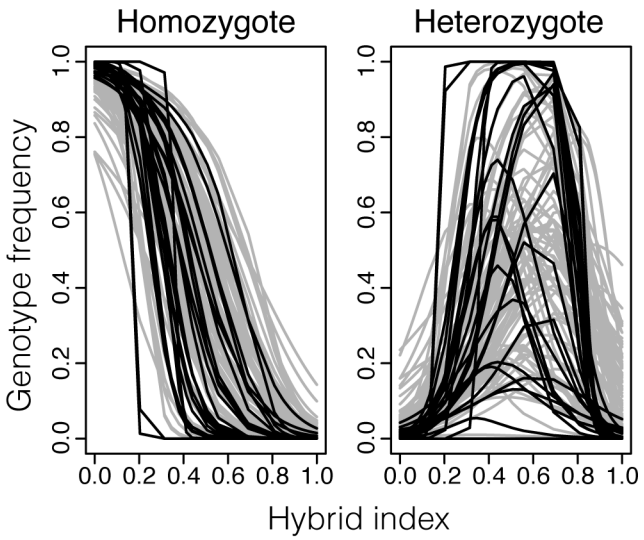


Figure 4.5 Summary of genomic clines for homozygous *G. pennsylvanicus* genotypes (left) and heterozygous *G. pennsylvanicus* and *G. firmus* genotypes (right). Each line represents the genomic cline for a single locus, plotted as the observed genotype class for each individual at that locus against the hybrid index. Black lines represent loci that significantly reduced introgression and grey lines are loci that either do not deviate from neutral introgression or have patterns of introgression consistent with overdominance or directional selection.

Table 4.4 Genotyping results of a representative SNP for each of the 15 genes encoding seminal fluid proteins assayed. Included are the genotyping call rate, allele frequencies (based on three allopatric populations of each species), interspecific differentiation index, deviation category from genomic cline analysis and functional homology.

Locus	**Call**	***Gp***	***Gf***	***D***	**Dev.**	**Functional homology**
AG-0383F	0.98	1	0.028	0.972	n.s.	Unknown
AG-0177F	1.00	1	1	0	-	Unknown
AG-0005F	1.00	0.186	0.861	0.675	-	Unknown
AG-0308F	0.99	1	1	0	-	Serine protease
AG-0115F	0.99	0.557	0.556	0.002	-	Unknown
AG-0159F	1.00	1	1	0	-	Trypsin protease
AG-0148P	1.00	1	0.125	0.875	P → F	Similar to lectin
AG-0203P	0.99	0.986	0.097	0.888	het +	Unknown
AG-0023F	0.00	-	-	-	-	Carboxipeptidase
AG-0254P	1.00	1	1	0	-	Chemiosensory protein
AG-0313F	0.99	0.671	0.403	0.269	-	Unknown
AG-0334P	0.99	1.000	0.361	0.639	-	Unknown
AG-0501F	1.00	0.014	1	0.986	n.s.	Proteasome
AG-0076F	0.40	-	-	-	-	Unknown
AG-0099F	1.00	0.500	0.681	0.181	-	Unknown

to rule a primer misalignment (due to repetitive sequence) that may have resulted in our genotype assay targeting the wrong nucleotide position. Across 232 SNPs assayed only nine were monomorphic and four of these were SFPs. Of the remaining nine markers, five had $D <$ 0.80. For the four markers with $D \geq 0.80$, AG-0148P had a pattern consistent with directional selection (*G. pennsylvanicus* allele advantageous); AG-0203P had an excess of heterozygotes consistent with overdominance and AG-0383F and AG-0501F did not deviate significantly from neutral expectations.

Discussion

Defining the genomic architecture of barriers to gene exchange is challenging, particularly in organisms with little genomic information, but there is the potential to gain insight into the diversity of mechanism and evolutionary process that contributes to speciation. Allopatric divergence and secondary contact of closely related species is a unique opportunity to identify genes that contribute to barriers between species. Genes identified through genome (or transcriptome) scans as having elevated divergence (high F_{ST} or major allele frequency differences) in otherwise undifferentiated genomes can be functionally tested for a role in barriers by estimating the extent of introgression; genomic regions with low gene flow between species reflect divergent selection, under-dominance, or assortative mating. Hybrid zones are essentially the ultimate crossing experiment conducted over numerous generations in natural settings, with recombination continually shuffling divergent genomes and selection determining the outcome of each cross (Harrison 1990; Payseur 2010). We identified 36 genome regions (contigs) that have significantly reduced gene flow within the field cricket hybrid zone. It is not clear how many of these regions are direct targets of selection or are simply linked to regions

that include “barrier genes”.

Patterns of differential introgression reveal candidate barrier genes

Genome scans have become a relatively easy way to identify outlier loci in comparisons of non-model taxa that have diverged relatively recently (e.g. Nosil et al. 2008; Manel et al. 2009; Schwarz et al. 2009; Apple et al. 2010; Galindo et al. 2010; Michel et al. 2010; Renaut et al. 2010; Fan et al. 2012; Nadeau et al. 2012; Stölting et al. 2012). However, genomic divergence can result from a number of processes, not all related to barriers to gene exchange or speciation. For instance, alternative alleles may become fixed in allopatry as a result of natural or sexual selection in different ecological or reproductive contexts, and although these traits may contribute to species differences, they may not be involved in the initial isolation or even maintenance of reproductive isolation. Favorable alleles may arise in a single lineage and introgress easily upon secondary contact. Heterogeneous divergence can also arise as a result of random lineage sorting in isolated populations or as a result of variable recombination rates across the genome (Noor and Bennett 2009). Indeed, we find that in our panel of 125 contigs that have elevated divergence in allopatric populations, the majority (71.2%) either introgress at the rate expected given the overall divergence of these two lineages or have elevated introgression. This is consistent with what is seen in other hybrid zones for which there are a large number of markers that show major allele frequency differences between allopatric populations (Teeter et al. 2008; Gompert et al. 2012a; Luttikhuizen et al. 2012). In these cases, many of the genes with the highest F_{ST} values outside of the hybrid zone did not necessarily exhibit reduced introgression in the hybrid zone.

The power of differential introgression analysis in hybrid zones is that it allows us to

narrow our focus to a subset of genome regions for which gene flow is less than expected. Many of the divergent genes we assayed (28.8%) have reduced introgression, consistent with patterns of disruptive selection, assortative mating or under-dominance. Functional annotation of these genes reveals several interesting classes of proteins that may play a role in barriers between *G. pennsylvanicus* and *G. firmus* (Table 4.3). The most abundant class of proteins we observed was cytoskeletal proteins that bind to actin or tubulin. These proteins are involved in several steps of fertilization, including sperm capacitation, the acrosome reaction, sperm-egg fusion and male and female pronuclei fusion (Fenichel and Durand-Clement 1998; Dvorakova et al. 2005; Sun and Schatten 2006; Sosnik et al. 2009). Divergence in proteins that mediate sperm-egg binding and fusion may be involved in the failure of *G. firmus* sperm to enter *G. pennsylvanicus* eggs (Larson et al. 2012b). The second most abundant class of proteins was growth hormone-stimulating proteins. These proteins could possibly be involved in the morphological differences in body-size and ovipositor length between *G. pennsylvanicus* and *G. firmus*. Both body-size and ovipositor length may play a role in ecological adaptation to different environments (Rand and Harrison 1989; Ross and Harrison 2006; Larson et al. 2013) and body-size has been implicated in mate choice in *G. firmus* (N Saleh, EL Larson, and RG Harrison unpublished data).

Is it surprising that we find so many loci that show patterns of disruptive selection? Similar numbers of markers with significantly reduced introgression have been identified in other hybrid zones (Lexer et al. 2007; Gompert and Buerkle 2009; Nolte et al. 2009; Teeter et al. 2010; Hamilton et al. 2012). However, patterns of differential introgression can result from multiple factors or forces, and distinguishing among these alternatives is difficult. Genetic drift, intraspecific population structure and incomplete lineage sorting can all result in heterogeneity in allele frequencies across populations. Field crickets have large effective populations sizes (2.6

million for *G. firmus* and 1.5 million for *G. pennsylvanicus*, Maroja et al. (2009a)) and are abundant throughout areas of suitable habitat, so neither genetic drift nor underlying population structure are likely to explain heterogeneity in patterns of introgression. Because we selected markers that had large differences in allele frequencies between allopatric populations, ancestral polymorphism also does not explain the variation in the extent of introgression. Sex-linked markers may have patterns of introgression that differ from the rest of the genome, because of differences in dispersal between males and females. Only one of the crickets with high levels of interspecific heterozygosity was male, so for many of our markers it is not possible to rule out sex-linkage (on the basis of identifying heterozygous males). But there is no evidence that male and female field crickets differ in their long-distance dispersal tendencies; both sexes can develop long hind-wings and are capable of flight during years of high abundance (Harrison 1979, 1980). Finally, there has been both empirical evidence and subsequent modeling that suggests that initial expectations about the ability of genomic cline modes to distinguish different modes of selection were overly optimistic (Macholan et al. 2011; Gompert et al. 2012b). These concerns are primarily in assigning deviation categories of epistasis, over-dominance and directional selection. These are instances where stochasticity in sampling the distribution of hybrid indices is more likely to lead to deviations in cline shapes, whereas a striking deficit of heterozygous genotypes is much less likely to be misinterpreted. As was noted by Dufková et al. (2011) careful marker selection and two-dimensional sampling of hybrid zones (as opposed to linear transects), as we have done here, makes differential introgression a reliable and useful tool for inferring selection on different genomic regions in natural populations.

The genomic regions we identify with restricted introgression across the hybrid zone appear to be excellent candidates for genes involved in barriers between species. The functional

annotation suggests several possible roles for the protein products of these genes. Without a linkage map, it is impossible to determine how many of these markers are the direct targets of selection versus how many are physically or genetically linked to targets of selection. Comparisons of associations between these markers suggest that there may be many associations. Knowledge of both the extent and composition of genomic divergence (so-called genomic islands of divergence, see Nosil and Feder (2012) and Harrison (2012)) and the distribution of these loci across the genome will provide insight the evolution of reproductive isolation between these species.

No evidence for reduced introgression of seminal fluid proteins

Genes encoding seminal fluid proteins (SFPs) have been proposed as potential candidate barriers genes between *G. pennsylvanicus* and *G. firmus* (Andrés et al. 2006; Braswell et al. 2006; Andrés et al. 2008; Maroja et al. 2009a). SFPs play a critical role in reproduction (reviewed in Wolfner 2009) and often have elevated divergence between closely related taxa (Swanson and Vacquier 2002; Clark et al. 2006; Turner and Hoekstra 2008). In *G. pennsylvanicus* and *G. firmus,* genes encoding SFPs have been shown to evolve rapidly and for some of these genes there is evidence of positive selection (Andrés et al. 2006; Andrés et al. 2008). Two cricket SFP genes, AG-0005F and AG-0334P, encode proteins with radical amino acid substitutions between species and have near zero introgression between allopatric populations (Maroja et al. 2009a). But comparisons between genes known to encode SFPs (identified through proteomic analyses of male ejaculate) and other genes expressed in the male accessory gland (the site of synthesis for SFPs) found that, as a class, SFPs did not show greater than expected differences in their average allele frequencies. In fact, many other genes

expressed in the male accessory gland are as differentiated or more differentiated than AG-0005F and AG-0334P (Andrés et al. 2013)

Here, we documented patterns of introgression within the hybrid zone for genes known to encode SFPs (see Andrés et al. 2008). We found that many of the SFPs we assayed did not have sufficient allele frequency differences to meet our criterion for genomic cline analysis. Indeed, in our much larger sample of individuals and populations, average allele frequency differences for AG-0005F and AG-0334P were < 0.8. Of the four SFP markers included in our introgression analysis (D > 0.8) one had a pattern of introgression consistent with directional selection (*G. pennsylvanicus* allele into *G. firmus* background), one had an excess of heterozygotes and two did not deviate significantly from neutral introgression.

In the past decade there has been an emphasis in the literature that genes encoding reproductive proteins, SFPs in particular, are among the most rapidly evolving genes in the genome, and this elevated divergence may contribute to reproductive isolation. We echo the caution raised by others that there is heterogeneity in the evolutionary rate of genes encoding SFPs, and although some may be rapidly evolving, a significant fraction may also be evolutionarily constrained (Andrés et al. 2006; Findlay et al. 2008; Dean et al. 2009; Walters and Harrison 2011; Andrés et al. 2013). This does not rule out a role for individual SFPs in post-mating prezygotic barriers. Indeed, the barriers we see between *G. pennsylvanicus* and *G. firmus* still strongly suggest that reproductive proteins may play a critical role in reproductive isolation (Larson et al. 2012a; Larson et al. 2012b). However, there is no evidence that the SFPs assayed here have a significant role in reducing gene exchange within the cricket hybrid zone.

Instead, we find evidence that other genes may be involved in post-mating prezygotic barriers. We identified one gene with reduced introgression within the hybrid zone that has a

functional class similar to previously described SFPs (peptidase inhibitor), but that has not been identified in proteomic analyses of the male ejaculate. We also identified several genes that are similar to those encoding cytoskeletal proteins that to bind to actin or tubulin, proteins that as a class have been implicated in several sperm/egg interactions. The role of these proteins in the post-mating prezygotic barriers between *G. firmus* and *G. pennsylvanicus* needs to be investigated further.

A bimodal hybrid zone maintained by strong prezygotic barriers

The composition of multi-locus genotypes within a hybrid provides insight into the forces maintaining hybrid zone structure. The distribution of genotypes can be used to infer the type of selection acting on individual loci (see above), but can also be used to look at the extent of hybridization and introgression within an ecological context and at the organismal level (Harrison 1990; Harrison and Bogdanowicz 1997; Jiggins and Mallet 2000).

In the cricket hybrid zone, there is a pattern of asymmetrical introgression for many of the markers assayed, with greater flow of *G. pennsylvanicus* alleles into *G. firmus* genomic backgrounds (Figure 4.4). For an individual marker, significantly greater introgression can represent the advancing wave of an advantageous allele that arose in a single lineage, but spreads into both lineages upon secondary contact (Barton and Hewitt 1985). However, directional selection does not explain a pattern of asymmetrical introgression across all markers. We find a pattern of greater introgression of *G. pennsylvanicus* alleles even at presumably neutral markers; there are a greater number of heterozygous genotypes in crickets that are characterized as *G. firmus* based on the hybrid index (Figure 4.4).

In other studies, we have seen evidence of greater flow of mtDNA into *G. firmus*, a pattern

consistent with the one-way incompatibility between *G. firmus* females and *G. pennsylvanicus* males. Because all hybrids carry *G. pennsylvanicus* mtDNA, backcrossing can only lead to movement of *G. pennsylvanicus* mtDNA haplotypes into *G. firmus* (Harrison et al. 1987; Harrison and Bogdanowicz 1997). Here, we also find that crickets that are predominantly *G. firmus* at nuclear makers (backcrosses into *G. firmus*) often carry *G. pennsylvanicus* mitochondrial DNA (crickets from sites I, K, U, BA, BB, AU, BX)(Larson et al. 2013). However, the one-way incompatibility does not explain the more extensive nuclear introgression into *G. firmus*. Hybrid offspring will contain nuclear alleles from each parental genome, and backcrossing should lead to introgression in both directions. Hybrid males appear unable to fertilize *G. firmus* females but hybrid females can backcross with either parent (Larson and Harrison unpublished data). In fact, mate-choice trials suggest that hybrid females prefer to mate with *G. firmus* males (Maroja et al. 2009b). Given these observations, alleles from either species should be incorporated into the genomic background of the other, perhaps at a slightly higher rate into *G. pennsylvanicus* given that both male and female hybrids can backcross with this species.

The observed asymmetry in presumably neutral makers could be a result of colonization history and differences in relative abundance in the hybrid zone. Areas where we see the greatest extent of introgression (Figure 4.2) are regions where *G. firmus* crickets appear to be expanding northward, up through the Great Appalachian valley, and moving into *G. pennsylvanicus* territory via roadways and mountain valleys. Wherever *G. firmus* was found in Pennsylvania, there were very high population densities (Larson et al. 2013). *Gryllus pennsylvanicus* may be the rarer species in areas of recent contact, and as a result, F1 hybrids would be more likely to backcross into *G. firmus*. Similar patterns of relative abundance

influencing the extent of hybridization and introgression have been seen in other hybrid zones, both in Chrysomelid beetles (Peterson et al. 2005) and in *Timema* walking sticks (Nosil et al. 2003). Asymmetry may also represent movement of the hybrid zone boundary, leaving a trail of neutral or weakly selected markers in its wake (Teeter et al. 2008; Teeter et al. 2010; Macholan et al. 2011).

The geographic distribution of hybrid indices corresponds to patterns we see for mtDNA and morphological traits; the hybrid zone in Pennsylvania is a mosaic of populations of each species (Figure 4.2)(Larson et al. 2013). In fact, morphological variation appears to be a generally good predictor of species identity. Although there is considerable overlap in morphological traits, multigenerational hybrids actually do have intermediate morphologies (Figure 4.3). Most of the sampled localities are predominantly *G. pennsylvanicus* or *G. firmus*, but also contain some backcross individuals (AK, which is pure *G. pennsylvanicus*, is an exception). These sites are patchily distributed, and we see abrupt transitions in allele frequencies between adjacent populations. Parental types are associated with different habitats; *G. pennsylvanicus* occupies habitat with more natural vegetation (forest clearings), while *G. firmus* occupies more disturbed habitat (agricultural and suburban areas) (Larson et al. 2013). The patchy distribution of habitat increases the opportunities for hybridization and introgression across patch boundaries and the dispersal of parental types into the other species habitat. There are a number of localities where both parental types (or backcrosses) co-occur (AR, BU, CE, BZ, Y, CI), but only in a few of these do we find F1s or multi-generation hybrids (J, L, AX, BK, BU, BV, CD).

The results of our admixture analysis are consistent with a hybrid zone that is maintained by strong barriers to gene exchange that prevent the formation of F1 hybrids and perpetuate

linkage disequilibrium. We see evidence of historical introgression over multiple generations (extensive backcrossing), but there is still nearly complete linkage disequilibrium among parental alleles and we find very few F1s or multi-generation hybrids (Figures 4.1 and 4.3). This signifies that barriers to gene exchange between these species are strong and substantially reduce the extent of hybridization in the face of ongoing opportunities for contact over multiple generations. Pre-zygotic barriers are well documented between these species and include temporal isolation (Harrison and Arnold 1982; Harrison 1985), habitat associations (Harrison 1986; Harrison and Rand 1989; Rand and Harrison 1989; Ross and Harrison 2006; Larson et al. 2013), behavioral isolation (Harrison and Rand 1989; Maroja et al. 2009b), and post-mating prezygotic barriers (Harrison 1983; Maroja et al. 2009b; Larson et al. 2012b). There is less evidence for post-zygotic barriers between these species. Crosses between *G. pennsylvanicus* females and *G. firmus* males produce viable, fertile hybrids and there is no evidence of exogenous selection against hybrids in natural settings. Crosses between *G. firmus* females and *G. pennsylvanicus* males produce almost entirely unfertilized eggs (with no evidence that the sperm enters the egg), but a small proportion of eggs (<3%) show signs of mitotic division (Larson et al. 2012b). In these cases it is impossible to rule out whether the observed mitotic division is a result of true fertilization and early embryonic death or if these eggs underwent activation and division of the female pronucleus (parthenogenesis). Thus, disruptive selection or assortative mating likely maintain linkage disequilibrium within the hybrid zone and explain the patterns of reduced gene flow we see for individual loci (see above). Many of the genes we have identified with reduced gene flow have functional annotation that suggests a role in prezygotic barriers (see above).

Decades of detailed analyses of hybrid zones have consistently documented patterns of

differential introgression. Early models focused on clinal hybrid zones characterized by sigmoidal transitions in parental allele frequencies across geographic space, with steep transitions occurring at the center of the hybrid zone. These models invoked a balance between dispersal and endogenous selection against hybrids (Key 1968; Barton and Hewitt 1985) or exogenous selection across an ecotone (Endler 1977) to maintain hybrid zone structure. But it was also recognized that environmental selection could reduce gene flow through pre-zygotic barriers, and underlying habitat heterogeneity could lead to a mosaic of parental types (Harrison and Rand 1989). Nongeographic differential introgression models were developed that could be applied to mosaic hybrid zones; these include Barton's concordance analysis (Szymura and Barton 1986) and more recently the genomic clines analysis of Gompert and Buerkle (2009) (but see Lexer et al. 2007 for an earlier version of this model).

The latter model has now been used on a handful of hybridizing taxa to identify genomic regions with reduced introgression (Lexer et al. 2007; Nolte et al. 2009; Teeter et al. 2010; Macholan et al. 2011; Janousek et al. 2012). In all of these cases, post-zygotic barriers appear to be more important in restricting gene flow. As a result, there has been a greater emphasis in the current literature on interpretation of these models in relation to postzygotic barriers (underdominance), and little mention of prezygotic barriers (assortative mating, disruptive selection) (but see Payseur 2010). Prezygotic barriers between recently diverged taxa are likely common, and may play a large role in the early stages of speciation (Harrison 1990; Coyne and Orr 2004; Shaw and Mullen 2011). The dynamics of hybrid zones maintained by prezygotic barriers may be very different than the dynamics of tension zones or clinal hybrid zones maintained by exogenous selection against hybrid genotypes. When interpreting the results of genomic cline analyses it is important to consider how prezygotic barriers influence patterns of

introgression.

Gryllus pennsylvanicus and *G. firmus* are very recently diverged, yet they are already at a stage in divergence where they are clearly distinguishable, with different morphology, ecology and behaviors. The bimodality we see in the distribution of hybrid indices suggests that they are on distinct evolutionary trajectories. Although there are multiple prezygotic barriers that isolate these linages, these barriers are incomplete. The boundary the separates these species is semipermeable, allowing some genomic regions to pass freely between species, while restricting those that are involved in reproductive barriers. Identifying genomic regions that fail to introgress is an important step towards understanding the genetic architecture of speciation and the evolutionary forces shaping divergence.

Acknowledgements

We thank ER Bondra for assistance with DNA extractions and L Cote at the Cornell Core Laboratories Center for Sequenom genotyping. This work was supported by a National Science Foundation grant DEB-0639904 to RGH and an AAUW American Fellowship to ELL.

References

Alexander, R. D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). Ann. Entomol. Soc. Am. 50:584-602.

Alexander, R. D. 1968. Life cycle origins speciation and related phenomena in crickets. Q. Rev. Biol. 43:1-41.

Andrés, J. A., E. L. Larson, S. M. Bogdanowicz, and R. G. Harrison. 2013. Assessing genomic differentiation in a pair of recently diverged field cricket species. Genetics in press.

Andrés, J. A., L. S. Maroja, S. M. Bogdanowicz, W. J. Swanson, and R. G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23:1574-1584.

Andrés, J. A., L. S. Maroja, and R. G. Harrison. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 275:1975-1983.

Apple, J. L., T. Grace, A. Joern, P. S. Amand, and S. M. Wisely. 2010. Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. Mol. Ecol. 19:4012-4028.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113-148.

Bazykin, A. D. 1969. Hypothetical mechanism of speciaton. Evolution 23:685-687.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Ser. B. (Stat. Method.) 57:289-300.

Braswell, W. E., J. A. Andres, L. S. Maroja, R. G. Harrison, D. J. Howard, and W. J. Swanson. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome 49:1069-1080.

Broughton, R. E., and R. G. Harrison. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163:1389-1401.

Buerkle, C. A., and L. H. Rieseberg. 2001. Low intraspecific variation for genomic isolation between hybridizing sunflower species. Evolution 55:684-691.

Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, R. F. C. Cajas, W. Diao, M. E. Maan, S. Paolucci, F. J. Weissing, L. van de Zande, A. Hoikkala, E. Geuverink, J. Jennings, M. Kankare, K. E. Knott, V. I. Tyukmaeva, C. Zoumadakis, M. G. Ritchie, D. Barker, E. Immonen, M. Kirkpatrick, M. Noor, C. M. Garcia, T. Schmitt, M. Schilthuizen, and S. N. Marie Curie. 2012. What do we need to know about speciation? Trends Ecol. Evol. 27:27-39.

Carling, M. D., and R. T. Brumfield. 2009. Speciation in *Passerina* buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. Mol. Ecol. 18:834-847.

Carling, M. D., R. T. Brumfield, and M. Webster. 2008. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. Evolution 62:2600-2615.

Clark, N. L., J. E. Aagaard, and W. J. Swanson. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131:11-22.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Dean, M. D., N. L. Clark, G. D. Findlay, R. C. Karn, X. Yi, W. J. Swanson, M. J. MacCoss, and M. W. Nachman. 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). Mol. Biol. Evol. 26:1733-1743.

Dufková, P., M. Macholán, and J. Piálek. 2011. Inference of selection and stochastic effects in the house mouse hybrid zone. Evolution 65:993-1010.

Dvorakova, K., H. D. M. Moore, N. Sebkova, and J. Palecek. 2005. Cytoskeleton localization in the sperm head prior to fertilization. Reproduction 130:61-69.

Endler, J. A. 1977. Geographic Variation, Speciation and Clines. Princeton University Press, Princeton, NJ.

Fan, S., K. R. Elmer, and A. Meyer. 2012. Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. Philos. Trans. R. Soc. B-Biol. Sci. 367:385-394.

Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil. 2012. Establishment of new mutations under divergence and genome hitchhiking. Philos. Trans. R. Soc. B-Biol. Sci. 367:461-474.

Fenichel, P., and M. Durand-Clement. 1998. Role of integrins during fertilization in mammals. Human Reproduction 13:31-46.

Findlay, G., X. Yi, M. MacCoss, and W. J. Swanson. 2008. Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS Biol. 6:1417-1426.

Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. J. Evol. Biol. 23:2004-2016.

Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol. Ecol. 18:1207-1224.

Gompert, Z., and C. A. Buerkle. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Mol. Ecol. Res. 10:378-384.

Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and C. A. Buerkle. 2012a. Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. Evolution 66:2167-2181.

Gompert, Z., T. L. Parchman, and C. A. Buerkle. 2012b. Genomics of isolation in hybrids. Philos. Trans. R. Soc. B-Biol. Sci. 367:439-450.

Hamilton, J. A., C. Lexer, and S. N. Aitken. 2012. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone. New Phytol. in press.

Harrison, R., and S. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36:535-552.

Harrison, R. G. 1979. Flight polymorphism in the field cricket *Gryllus pennsylvanicus*. Oecologia 40:125-132.

Harrison, R. G. 1980. Dispersal polymorphisms in insects. Annu. Rev. Ecol. Syst. 11:95-118.

Harrison, R. G. 1983. Barriers to gene exchange between closely related cricket species . I. laboratory hybridization studies. Evolution 37:245-251.

Harrison, R. G. 1985. Barriers to gene exchange between closely related cricket species . II. life-cycle variation and temporal isolation. Evolution 39:244-259.

Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56:337-349.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Pp. 69-128 *in* D. Futuyma, and J. Antonovics, eds. Oxford Surveys in Evolutionary Biology. Oxford University Press, New York, NY.

Harrison, R. G. 1991. Molecular changes at speciation. Annu. Rev. Ecol. Syst. 22:281-308.

Harrison, R. G. 1993. Hybrids and hybrid zones: historical perspective. Pp. 3-12 *in* R. G. Harrison, ed. Hybrid Zones and the Evolutionary Process. Oxford University Press, Oxford.

Harrison, R. G. 2010. Understanding the origin of species: where have we been, where are we going? Pp. 688 *in* M. A. Bell, D. J. Futuyma, W. F. Eanes, and J. S. Levinton, eds. Evolution since Darwin: The First 150 Years. Sinauer Associates.

Harrison, R. G. 2012. The language of speciation. Evolution DOI: 10.1111/j.1558-5646.2012.01785.x.

Harrison, R. G., and S. M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51:493-505.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 111 - 133 *in* D. Otte, and J. A. Endler, eds. Speciation and its Consequences. Sinauer, Sunderland, MA.

Harrison, R. G., D. M. Rand, and W. C. Wheeler. 1987. Mitochondrial-DNA variation in field crickets across a narrow hybrid zone. Mol. Biol. Evol. 4:144-158.

Hewitt, G. M. 1988. Hybrid zones - natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158-167.

Janousek, V., L. Y. Wang, K. Luzynski, P. Dufkova, M. M. Vyskocilova, M. W. Nachman, P. Munclinger, M. Macholan, J. Pialek, and P. K. Tucker. 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. Mol. Ecol. 21:3032-3047.

Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends Ecol. Evol. 15:250-255.

Key, K. H. L. 1968. The concept of stasipatric speciation. Syst. Zool. 17:14-22.

Larson, E. L., J. A. Andrés, and R. G. Harrison. 2012a. Influence of the male ejaculate on post-mating prezygotic barriers in field crickets. PLoS ONE 7:e46202.

Larson, E. L., G. C. Becker, E. R. Bondra, and R. G. Harrison. 2013. Structure of a mosaic hybrid zone between the field crickets *G. firmus* and *G. pennsylvanicus*. Ecology and Evolution In review.

Larson, E. L., G. L. Hume, J. A. Andrés, and R. G. Harrison. 2012b. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. J. Evol. Biol. 25:174-186.

Lexer, C., C. A. Buerkle, J. A. Joseph, B. Heinze, and M. F. Fay. 2007. Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. Heredity 98:74-84.

Lexer, C., J. A. Joseph, M. van Loo, T. Barbara, B. Heinze, D. Bartha, S. Castiglione, M. F. Fay, and C. A. Buerkle. 2010. Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. Genetics 186:699-712.

Luttikhuizen, P. C., J. Drent, K. Peijnenburg, H. W. van der Veer, and K. Johannesson. 2012. Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica*. Mol. Ecol. 21:3048-3061.

Macholan, M., S. J. E. Baird, P. Dufkova, P. Munclinger, B. V. Bimova, and J. Pialek. 2011. Assessming multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. Evolution 65:1428-1446.

Manel, S., C. Conord, and L. Despres. 2009. Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. BMC Evol. Biol. 9:288.

Maroja, L. S., J. A. Andrés, and R. G. Harrison. 2009a. Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. Evolution 63:2999-3015.

Maroja, L. S., J. A. Andrés, J. R. Walters, and R. G. Harrison. 2009b. Multiple barriers to gene exchange in a field cricket hybrid zone. Biol. J. Linn. Soc. 97:390-402.

Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. Proc. Natl. Acad. Sci. USA 107:9724-9729.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. Philos. Trans. R. Soc. B-Biol. Sci. 367:343-353.

Nolte, A. W., Z. Gompert, and C. A. Buerkle. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. Mol. Ecol. 18:2615-2627.

Noor, M. A. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439-444.

Noor, M. A. F., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7:851-861.

Nosil, P., B. J. Crespi, and C. P. Sandoval. 2003. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. Proceedings of the Royal Society of London Series B-Biological Sciences 270:1911-1918.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316-336.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B-Biol. Sci. 367:332-342.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375-402.

Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160-167.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol. Ecol. Res. 10:806-820.

Peterson, M. A., B. M. Honchak, S. E. Locke, T. E. Beeman, J. Mendoza, J. Green, K. J. Buckingham, M. A. White, and K. J. Monsen. 2005. Relative abundance and the species-specific reinforcement of male mating preference in the *Chrysochus* (Coleoptera : Chrysomelidae) hybrid zone. Evolution 59:2639-2655.

Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone - mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432-449.

Renaut, S., A. W. Nolte, and L. Bernatchez. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Mol. Ecol. 19:115-131.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713-727.

Ross, C. L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56:2296-2312.

Ross, C. L., and R. G. Harrison. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115:53-68.

Schwarz, D., H. M. Robertson, J. L. Feder, K. Varala, M. E. Hudson, G. J. Ragland, D. A. Hahn, and S. H. Berlocher. 2009. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. BMC Genomics 10:633.

Shaw, K., and S. Mullen. 2011. Genes versus phenotypes in the study of speciation. Genetica 139:649-661.

Sosnik, J., P. V. Miranda, N. A. Spiridonov, S.-Y. Yoon, R. A. Fissore, G. R. Johnson, and P. E. Visconti. 2009. Tssk6 is required for Izumo relocalization and gamete fusion in the mouse. J. Cell Sci. 122:2741-2749.

Stölting, K. N., R. Nipper, D. Lindtke, C. Caseys, S. Waeber, S. Castiglione, and C. Lexer. 2012. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Mol. Ecol. in press.

Sun, Q.-Y., and H. Schatten. 2006. Regulation of dynamic events by microfilaments during oocyte maturation and fertilization. Reproduction 131:193-205.

Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3:137-144.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution 40:1141-1159.

Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien, J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman, and P. K. Tucker. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18:67-76.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker. 2010. The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64:472-485.

Turner, L. M., and H. E. Hoekstra. 2008. Causes and consequences of the evolution of reproductive proteins. Int. J. Dev. Biol. 52:769-780.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:1572-1578.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. B-Biol. Sci. 367:451-460.

Walters, J. R., and R. G. Harrison. 2011. Decoupling of rapid and adaptive evolution among seminal fluid proteins in *Heliconius* butterflies with divergent mating systems. Evolution 65:2855-2871.

Willett, C. S., M. J. Ford, and R. G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79:484-494.

Wolfner, M. F. 2009. Battle and ballet: molecular interactions between the sexes in *Drosophila*. J. Hered. 100:399-410.

Wu, C. I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851-865.

CHAPTER 5

GENE FLOW AND THE MAINTENANCE OF SPECIES BOUNDARIES IN A MOSAIC HYBRID ZONE

Abstract

Patterns of differential introgression within hybrid zones can be used to evaluate directly the role of genomic regions in reproductive isolation between diverging lineages. Comparisons of different geographic, topographic, and ecological settings are critical to interpreting patterns of introgression. Here, we use estimates of locus-specific introgression for a large genomic dataset to identify genomic regions that may contribute to reproductive barriers between the field crickets *Gryllus pennsylvanicus* and *G. firmus*. We compare patterns of introgression between two regions of the field cricket hybrid zone sampled at two very different spatial scales. We find remarkably consistent patterns of introgression for individual loci, despite much greater overall introgression at the fine spatial scale. Genes with reduced introgression in both regions are clearly under selection in the hybrid zone.

Introduction

Species are often viewed as discrete entities, but the origin of species is an extended process of divergence and the evolution of reproductive isolation. Divergence and reproductive isolation are not characteristics of whole organisms (or lineages), but of individual genes or genome regions. Species share ancestral alleles for many genes across the genome, whereas for other genes, species have unique alleles that have diverged as a result of either random or selective lineage sorting (Harrison 1986, 1990; Wu 2001; Nosil et al. 2009). Genes that contribute to barriers between species must reside in these divergent genomic regions, and although their evolution is a by-product of the processes that drive divergence, it is the evolution of barriers to gene exchange that distinguishes speciation from intraspecific divergence. A detailed understanding of both the extent of genomic divergence between recently diverged

species and the genetic architecture of reproductive isolation (number, effect size, and chromosomal distribution of genes) is critical to elucidating the evolutionary processes that lead to speciation. New technologies for obtaining sequence data have made it relatively simple to exploit genomic heterogeneity to identify divergent genomic regions between closely related taxa (e.g. Nosil et al. 2008; Wood et al. 2008; Manel et al. 2009; Schwarz et al. 2009; Apple et al. 2010; Galindo et al. 2010; Renaut et al. 2010; Fan et al. 2012; Nadeau et al. 2012; Stölting et al. 2012; Andrés et al. 2013). A common approach is to look for correlations between divergent genomic regions (F_{ST} outliers) and species-specific traits that may contribute to reproductive isolation. However, the relationship between divergence and reproductive isolation is not so straightforward (see Gompert et al. 2012; Luttikhuizen et al. 2012). Divergence can result from a number of processes other than selection, such as genetic hitchhiking, drift, and variation in recombination rates (Noor and Bennett 2009; Nosil et al. 2009; Michel et al. 2010; Nosil and Feder 2012), or selection may lead to adaptive divergence in species traits that are unrelated to barriers.

Hybrid zones provide a unique opportunity to evaluate directly the role of genomic regions in reproductive isolation (Barton and Hewitt 1985; Hewitt 1988; Harrison 1990; Payseur 2010). Hybridization and introgression result in the recombination and shuffling of divergent genomes. Divergent gene regions that are unrelated to barriers between species (e.g., equally fit in either genomic background) will be easily exchanged between hybridizing taxa, whereas genes that are involved in barriers will show limited or no introgression. We can compare the strength of selection on different genomic regions by comparing the slopes for the change in allele or genotype frequencies across a gradient of hybridization (i.e., across a geographic transect or as a function of hybrid index) (Barton and Hewitt 1985; Szymura and Barton 1986;

Gompert and Buerkle 2009). Alleles that are neutral with respect to barriers will exhibit relatively shallow clines, while alleles at barrier genes will have characteristically steep clines (Key 1968; Bazykin 1969; Barton and Hewitt 1985).

For decades hybrid zone studies have been taking this approach to understanding the genetic architecture of species boundaries (reviewed in Harrison 1990; Payseur 2010). But we now have the opportunity to leverage large sequencing and genotyping datasets in natural populations to look at the extent of introgression on a genome-wide scale (e.g. Gompert et al. 2012; Janousek et al. 2012). There are a handful of systems that have taken this multi-locus approach (Lexer et al. 2007; Nolte et al. 2009; Teeter et al. 2010; Macholan et al. 2011; Gompert et al. 2012; Hamilton et al. 2012; Janousek et al. 2012; Larson et al. 2013a). Yet, there are still only a few organisms for which large genomic datasets have been used to compare introgression across different regions of the hybrid zone (Nolte et al. 2009; Teeter et al. 2010). Given that hybrid zone structure can vary dramatically across broad areas of contact (e.g. *Bombina* hybrid zone: Szymura and Barton 1991; Vines et al. 2003; Yanchukov et al. 2006), comparisons of patterns of introgression in different geographic, topographic, and ecological settings is critical to interpreting the genetic architecture of reproductive isolation. Spatial scale is an important component in interpreting patterns of variation across different regions. For instance, patchy hybrid zones could be perceived as having different patterns of introgression if sampled at a broad spatial scale, especially along one-dimensional transects (Dufková et al. 2011; Macholan et al. 2011). At a broad spatial scale, mosaic hybrid zones may appear clinal, but finer sampling reveals a patchwork of parental and mixed populations in close proximity (Harrison and Rand 1989). Hybridization and introgression occur across local patch boundaries, which may be the

most appropriate sampling scale (Bridle et al. 2002; Ross and Harrison 2002). Spatial scale is likely important in many hybrid zones, but is seldom addressed explicitly.

In this study, we focus on a well-characterized hybrid zone between two closely related field crickets, *Gryllus pennsylvanicus* and *G. firmus*, which are estimated to have diverged 200,000 years ago (Willett et. al. 1997; Broughton and Harrison 200x; Maroja et al. 2009). The hybrid zone stretches from Massachusetts to Virginia along the eastern edge of the Appalachian Mountains (Harrison and Arnold 1982; Larson et al. 2013b). *Gryllus pennsylvanicus* occupies the inland/upland areas to the west and north and *G. firmus* occupies the lowland/coastal areas to the east and south (Alexander 1957; Harrison and Arnold 1982). The hybrid zone appears clinal at a broad geographic scale (Harrison and Arnold 1982), but at intermediate scales there is a patchy distribution of pure individuals of each species, characteristic of a mosaic hybrid zone (Harrison 1986; Harrison and Rand 1989; Rand and Harrison 1989; Larson et al. 2013b). Hybridization and introgression occur at the boundaries of these parental populations and on a very fine scale there is a clinal transition in allele frequencies (Ross and Harrison 2002).

Hybrid zone structure appears to be maintained by underlying habitat heterogeneity, but the nature of habitat patches varies in different regions of the hybrid zone. In Connecticut, the advance and retreat of glacial ice has left a patchy distribution of soil type; ridges running perpendicular to the coast are characterized by loamy soils occupied by *G. pennsylvanicus*, while the intervening river drainages are primarily sandy and inhabited by *G. firmus*. Out-pockets of loamy upland-like habitat stretch to the coast while sandy soils extend inland along the river valleys, and the species interact at the boundaries of these patches (Harrison and Rand 1989; Rand and Harrison 1989; Ross and Harrison 2002). In Pennsylvania, a patchy distribution of natural habitat along mountain ridges is occupied by *G. pennsylvanicus*, while the primarily

agricultural and suburban lowlands are occupied by *G. firmus* (Larson et. al. 2013). The two species come into contact along corridors of disturbed habitat through the mountains and small pockets of natural habitat in the valleys and along mountain slopes.

In areas where the two species co-occur, there is evidence of assortative mating and post-mating prezygotic barriers. *Gryllus pennsylvanicus* females from mixed populations produce offspring sired primarily by conspecific males (Harrison 1986), but there is no evidence of conspecific sperm precedence (Larson et al. 2012b). In laboratory mate choice trials, females of both species are reluctant to mate with heterospecific males (Maroja et al. 2009b), suggesting that behavioral cues contribute to assortative mating. In crosses between *G. firmus* females and *G. pennsylvanicus* males, there is a one-way incompatibility; females never produce offspring (Harrison 1983). The incompatibility is a result of a combination of post-mating prezygotic barriers that reduce oviposition (Maroja et al. 2009b) and prevent stored *G. pennsylvanicus* sperm from properly fusing with *G. firmus* eggs (Larson et al. 2012b). In contrast, the reciprocal cross (*G. pennsylvanicus* female and *G. firmus* male) produces viable fertile offspring.

Here, we use estimates of locus-specific introgression for a large panel of single nucleotide polymorphism (SNP) markers to identify genomic regions that may contribute to these reproductive barriers. Genomic regions that have restricted gene flow within the hybrid zone reflect assortative mating, disruptive selection or underdominance. We compare patterns of introgression between two regions of the field cricket hybrid zone sampled at two very different spatial scales; an intermediate scale across a patchy distribution of parental and mixed populations (Pennsylvania) and a fine scale across a single patch boundary (Connecticut). We find remarkably consistent patterns of introgression for individual loci, despite much greater

overall introgression at the fine spatial scale. Genes with reduced introgression in both regions are clearly under selection in the hybrid zone.

Materials and Methods

Sampling

Reference populations- We used a previously published dataset of cricket genotypes from three allopatric populations of each species (11-12 individuals per population) to estimate the allele frequencies within each species (Larson et al. 2013a). The dataset included 35 *G. pennsylvanicus* from Ithaca, NY (ITH, N: 42°26'01", W: 76°29'59"); Scranton, PA (SCR, N: 41°24'25", W: 75°35'46"); and State College, PA (SCO, N: 40°47'59",W: 77°52'05") and 36 *G. firmus* from Guilford, CT (GUI, N: 41°16'48", W: 72°42'02"); Tom's River, NJ (TRI, N: 39°45'00", W: 74°11'33"); and Parksley, VA (PAR, N: 37°45'58", W: 75°36'00").

Connecticut- Crickets from the Connecticut hybrid zone were collected in September of 1996 and 1997 along River Road near the Connecticut River southeast of Middleton, CT (N: 41° 33'30"; W: 72°35'18"). The sampling area was a 500 m stretch that spanned a boundary between loam (0 m, *G. pennsylvanicus*) and sand (500 m, *G. firmus*) soil patches. Crickets were collected along both sides of the road and in surrounding habitat up to 20 m from the road. Collecting sites are described in Ross and Harrison (2002). Live crickets from both transects were brought back to the laboratory and frozen at -80°C. The majority of crickets were collected as adults, but in a few cases, crickets were collected as late instar nymphs. Nymphs were allowed to mature in the laboratory before freezing. We genotyped a total of 260 crickets; 68 crickets from loam, 46 crickets from sand and 146 from the transition between loam and sand patches.

Pennsylvania- We re-analyzed a previously published dataset of cricket genotypes from the Pennsylvania hybrid zone described in Larson et al. (2013a). These crickets were collected in August and September of 2008 and 2010 from south-central Pennsylvania. The sampling area spanned approximately 200 km^2 across the transition from the folds of the Appalachian Mountains into the Great Appalachian Valley and coastal plains. The dataset comprised a total of 301 crickets from 36 localities from the center of the hybrid zone where there was greatest extent of admixture estimated from distribution of mitochondrial haplotypes and morphological variation described in Larson et al. (2013b).

Molecular markers

We extracted genomic DNA from crickets collected in the Connecticut region using single adult femurs, thoraces or heads (depending on the available tissue). The majority of the crickets were extracted using a DNeasy Tissue Kit (QIAGEN Inc., Valencia, CA), but a subset of samples was extracted using phenol/chloroform (Ross and Harrison 2002). All DNA was diluted to 10 ng/ul. We genotyped Connecticut crickets using the MassARRAY platform (Sequenom Inc., San Diego, CA) for a subset of 151 bi-allelic SNPs originally described in Larson et al (Larson et al. 2013a). The MassARRAY platform uses a multiplexed amplification of target DNA followed by a single base extension of a primer immediately adjacent to the target SNP. The extend product was analyzed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) and differences in mass between each possible nucleotide allowed high quality genotyping. Reactions were performed using iPLEX Gold chemistry at the Cornell Life Sciences Core Laboratories Center for Genomics. We called SNP genotypes using the Sequenom MassARRAY Typer v4.0 Analysis software and confirmed genotype calls by eye.

We restricted our analyses to those SNPs with large allele frequency differences (≥ 0.80) between species (based on frequencies in six allopatric populations). We defined the interspecific differentiation index (*D*):

$$D = \left| P_{Gf} - P_{Gp} \right|,$$

(see Renaut et al. 2010; Andrés et al. 2013). Three SNPs described in Larson et al. (Larson et al. 2013a) were left out of the analyses because they had poor amplification (amplified in < 95% of individuals) in the new assay mixes. In total we included in our analyses genotypes from 301 Pennsylvania crickets and 260 Connecticut crickets for 114 SNPs.

Admixture and genomic clines analyses

We used the R-package INTROGRESS (Gompert and Buerkle 2009; Gompert and Buerkle 2010) to quantify genomic admixture and estimate genomic clines for both regions of the hybrid zone. We quantified admixture for each cricket by calculating the hybrid index, defined as proportion of alleles from all 114 markers that were inherited from *G. firmus*, using the function *'prepare.data'* (hybrid index: 0 = *G. pennsylvanicus*, 1 = *G. firmus*). We estimated the interspecific heterozygosity, defined as the proportion of alleles inherited from both parental populations using the function *'calc.intsp.het'*. Following Milne and Abbott (2008), crickets were assigned to hybrid classes based on their estimated interspecific heterozygosity and hybrid index. Crickets with a high interspecific heterozygosity (≥ 85%) and an intermediate hybrid index (~ 0.5) were classified as first generation hybrids (F1). Crickets with an interspecific heterozygosity < 85% were classified as either multi-generation hybrids (hybrid index between ≥ 0.25 and ≤ 0.75) or backcross into *G. pennsylvanicus* and *G. firmus* (hybrid index < 0.25 or > 0.75, respectively).

Genomic clines were estimated using multinomial regression to predict, based on the hybrid index and interspecific heterozygosity, the probability of observing each of the possible genotypes (PP: homozygous *G. pennsylvanicus*; PF: heterozygous; and FF: homozygous *G. firmus*) at each marker. To identify markers that do not conform to expectations of neutral introgression, the likelihood of the regression model was compared to that expected under a neutral model of introgression constructed from 2,000 parametric simulations based on the observed genotype frequencies, with the assumption that alleles at each marker are co-dominant. For all analyses, estimates of the hybrid index and interspecific heterozygosity were calculated based on the allele frequencies in the parental populations, and significance thresholds were adjusted using the false discovery rate procedure (Benjamini and Hochberg 1995). Evidence for an excess (+) or deficiency (-) of either homozygous or heterozygous genotypes was based on the proportion of neutral simulations that yield a model with higher total probability of a given genotype than the model based on the observed data. We summarized these deviations as either 1) excess or deficit of homozygotes (direct selection) 2) excess of heterozygotes (overdominance) or 3) deficit of heterozygotes (assortative mating, disruptive selection or underdominance) using a combination of the INTROGRESS output and visual inspection of cline shape.

Geographic clines

Geographic clines were estimated for the Connecticut region of the hybrid zone based on the methods developed by Szymura and Barton (Szymura and Barton 1986; Szymura and Barton 1991). We fitted simple two-parameter model tanh clines for each locus in R (R Core Development Team, 2010) using the methods described in Polyakov et al. (2011). Allele

frequencies were entered into a generalized linear model (GLM) with a logit link function and a binomial error structure, with distance along the linear transect as an explanatory variable. The center of the cline was calculated using the '*dose.p*' function available in the MASS package (Venables and Ripley 2002) and cline width was calculated as $w = 4/s$, where s is the logit slope parameter (Gay et al. 2008).

Results

Genomic structure of the hybrid zone

We scored 114 bi-allelic SNP markers in crickets from within the Pennsylvania (N = 301) and Connecticut (N = 260) regions of the hybrid zone (see Figure 5.1). For crickets in each region, we plotted the hybrid index for each individual against its interspecific heterozygosity, the proportion of loci at which that individual was heterozygous for alleles from the two parental species (Figure 5.2). At a regional scale in Pennsylvania (sampling ~ every 2-5 km over a 200 km^2 region), we found 65 pure *G. pennsylvanicus* and 30 pure *G. firmus*, but most crickets were classified as backcrosses into *G. pennsylvanicus* (128) and *G. firmus* (137). We found one cricket that could be considered an F1 hybrid and eight that are multi-generation hybrids. At a much smaller spatial scale in Connecticut (a transect of <1km), we found no pure *G. pennsylvanicus* crickets and eleven pure *G. firmus* crickets. The majority of crickets were backcrosses into *G. pennsylvanicus* (157) or *G. firmus* (96), and seven were multi-generation hybrids. Overall, there is more introgression in the Connecticut region of the hybrid zone (fewer pure parental types and more backcrosses) and there is greater variance in the hybrid indices among *G. pennsylvanicus*-like crickets. The spatial distribution of *G. pennsylvanicus* (orange), *G. firmus* (green) and hybrid classes are shown in Figure 5.1. In Pennsylvania, collecting

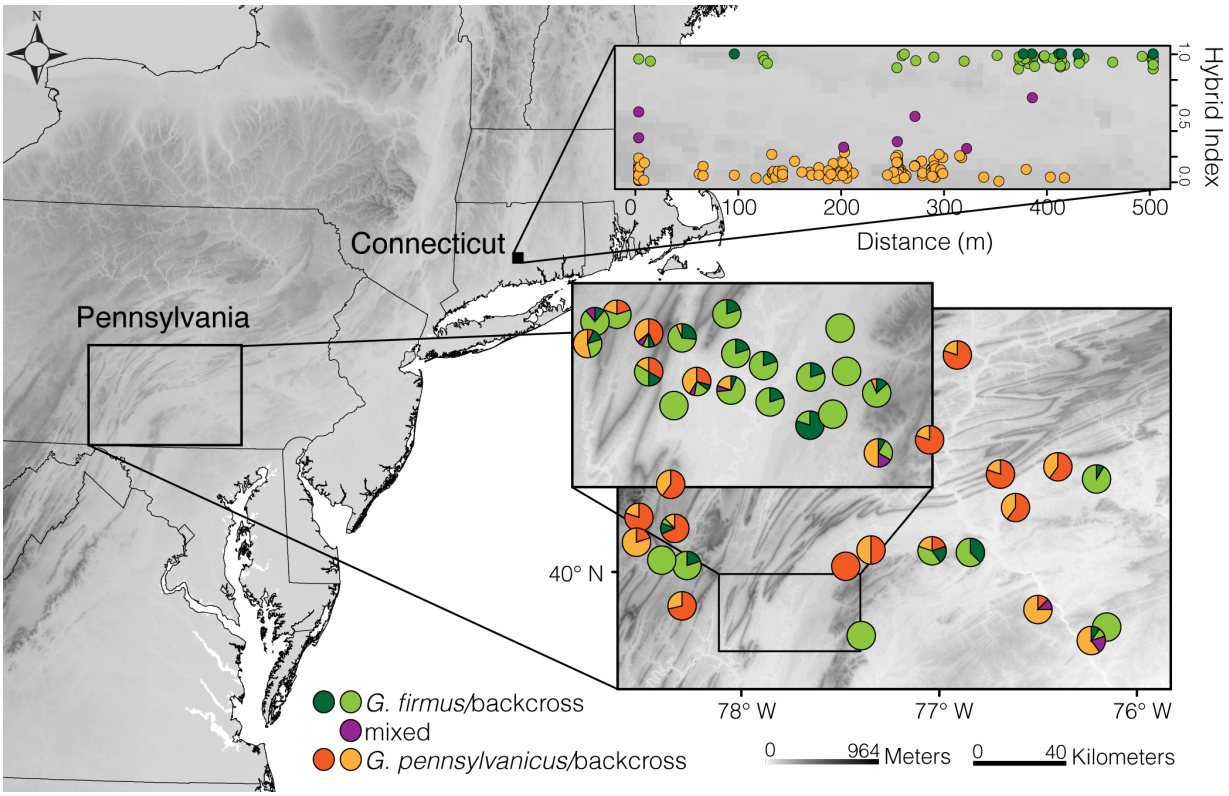


Figure 5.1 Maps of study areas in Pennsylvania (~ 200 km^2) and Connecticut (500 m) and the distribution of hybrid classes; *G. pennsylvanicus* (dark orange), *G. firmus* (dark green), backcrosses in to *G. pennsylvanicus* (light orange) and *G. firmus* (light green) and multi-generation hybrids (purple). In Pennsylvania, the pie charts represent the proportion of hybrid classes at each collecting locality and in Connecticut each cricket's hybrid index is color coded based on hybrid class and plotted against transect distance (m) from the *G. pennsylvanicus* end (0 m) on loamy soil to the *G. firmus* end (500 m) on sandy soil (note: the transect runs north to south).

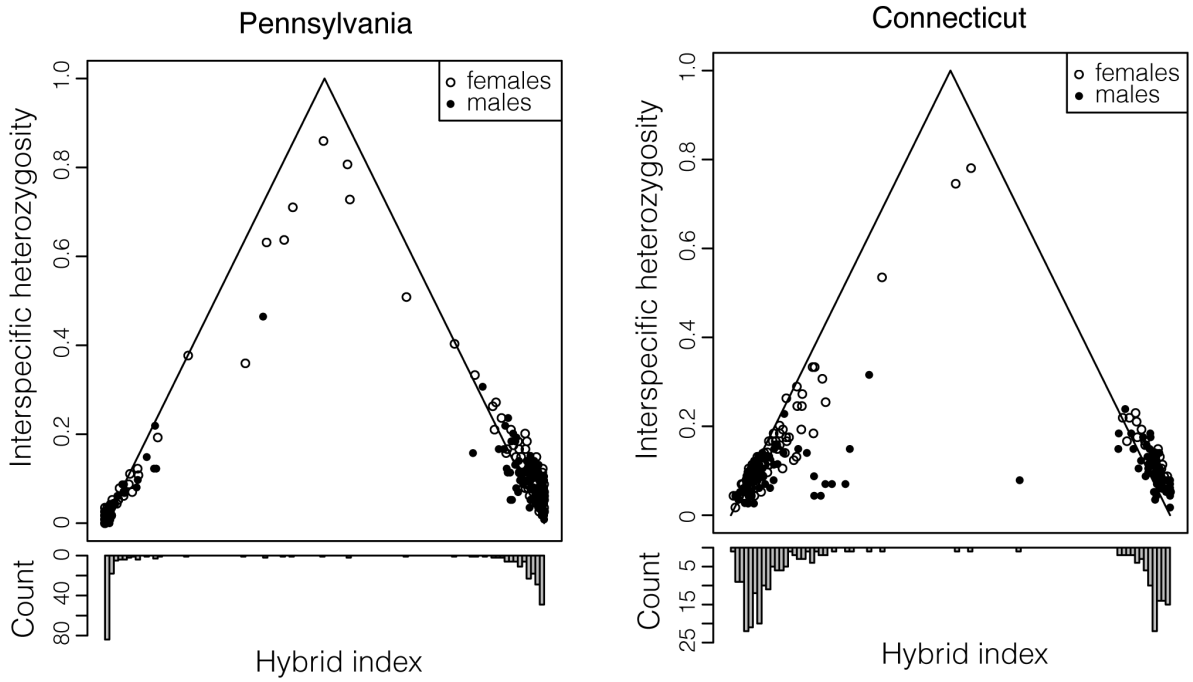


Figure 5.2 Interspecific heterozygosity plotted against the hybrid index for females (open circles) and males (closed circles) in the Pennsylvania and Connecticut regions of the hybrid zone. The hybrid index is measured as the proportion of alleles with *G. firmus* ancestry; a hybrid index of zero denotes pure *G. pennsylvanicus* and a hybrid index of one denotes pure *G. firmus*. The distribution of hybrid indices is plotted below. Crickets with high interspecific heterozygosity (≥ 85%) and intermediate hybrid index (~ 0.5) are F1 hybrids, while hybrid indices between 0.25 and 0.75 are multi-generation hybrids. Crickets with hybrid indices > 0.25 or < 0.75 are backcrosses into *G. pennsylvanicus* and *G. firmus* (respectively).

localities are predominantly composed of pure parental types (dark green and orange) or backcrosses (light green and orange). There are only eleven localities where we find both parental types, and only seven where we see F1s or multi-generation hybrids. In Connecticut, there is an abrupt transition between *G. pennsylvanicus*-like crickets and *G. firmus*-like cricket at approximately 300 – 400 m, but multi-generation hybrids are found both in the middle of the transect and at the *G. pennsylvanicus* end.

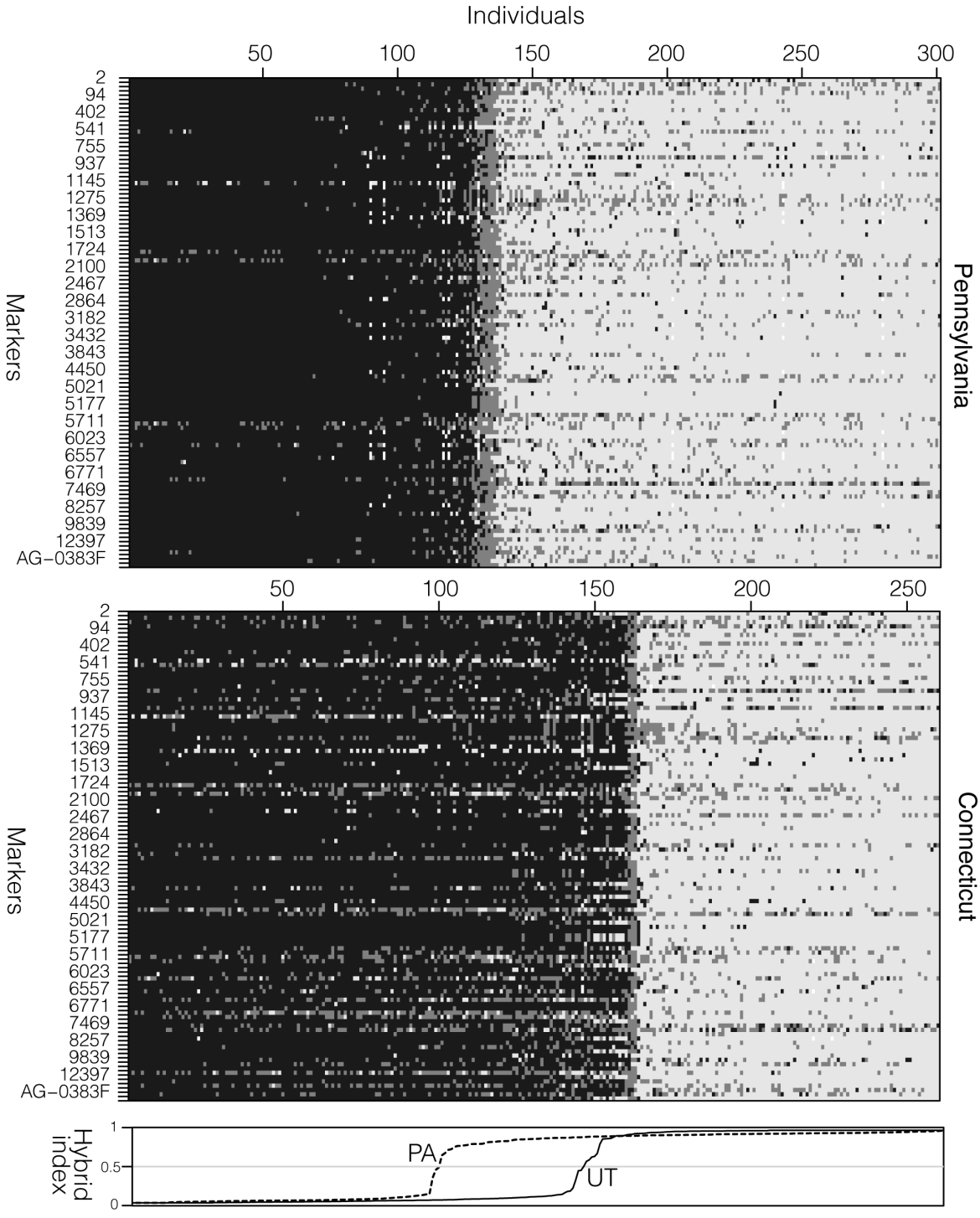
Genomic clines

The extent of introgression varied substantially among individual markers (Figure 5.3 and Figure S5.1), but patterns for each marker were consistent between the two regions of the hybrid zone (Table S5.1). There were 33 markers that had restricted introgression in Pennsylvania, and all of these also had restricted introgression in Connecticut (solid black lines Figure 5.4A). Eighteen additional markers also showed restricted introgression in Connecticut (dashed black lines Figure 5.4A), but these markers did not deviate significantly from neutrality in Pennsylvania. However, in Pennsylvania observed genotype frequencies and cline shapes are consistent with the patterns seen in Connecticut (Figure S5.1). Overall, the majority of the markers that are "different' between the two hybrid zones are those that do not deviate from neutrality in Pennsylvania, but do in Connecticut (Table 5.1). Only ten markers had inconsistent patterns between the two regions (Table S5.1).

Geographic clines

Geographic cline widths ranged from 167.6 to 842.3 m with centers between 123.6 and 588.6 m (Table S5.2). Markers that have significantly restricted introgression (based on genomic

Figure 5.3 Overview of patterns of introgression for 114 SNP markers genotyped both Pennsylvania (N= 301) and Connecticut (N = 260). Each row represents a single marker (every third marker is labeled) and each column represent an individual's genotype at that locus. Dark grey is homozygous for the *G. pennsylvanicus* allele, medium grey is heterozygous, light grey is homozygous for the *G. firmus* allele and white represents missing data. Below is a visual representation of the hybrid index: the proportion of each individual's genome that has *G. firmus* ancestry (*G. pennsylvanicus*: 0; *G. firmus*: 1).


Individuals
50
100
150
200
250
300
Markers
2
94
402
541
755
937
1145
1275
1369
1513
1724
2100
2467
2864
3182
3432
3843
4450
5021
5177
5711
6023
6557
6771
7469
8257
9839
12397
AG−0383F
Pennsylvania
50
100
150
200
250
Markers
Connecticut
Hybrid index
1
0.5
0
PA
UT

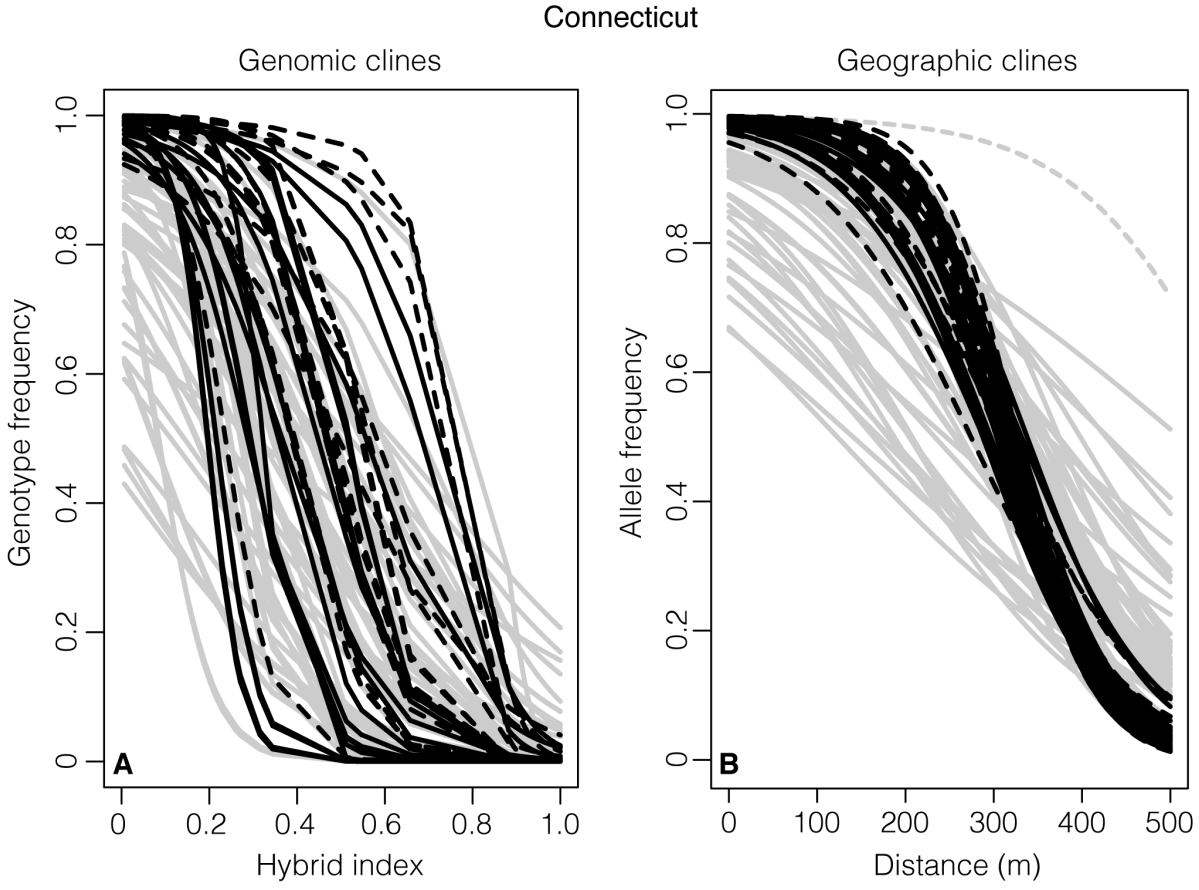


Figure 5.4 Plots of the genomic (A) and geographic (B) clines for 114 SNP markers in the Connecticut study area. Genomic clines are drawn based on the observed frequency of the *G. pennsylvanicus* genotype plotted against the hybrid index. Geographic clines are drawn as the frequency of the *G. pennsylvanicus* allele plotted against transect distance (m). In both panels, black lines represent markers with significantly reduced introgression in both the Pennsylvania and Connecticut regions and dashed lines represent markers that have significantly reduced introgression in Connecticut. Grey lines represent markers that either do not deviate from neutral expectations or have patterns of introgression consistent with directional selection or overdominance. The dashed grey line in panel B represents mtDNA.

Table 5.1 Summary of patterns of introgression for 114 markers genotyped in both the Pennsylvania and Connecticut regions of the hybrid zone. Markers exhibiting the same or different patterns of introgression between the hybrid zones are tallied. The majority of the makers that deviated from neutrality in one region but not the other, exhibited similar patterns of genotype distribution.

Pattern of introgression	**PA**	**CT**	**Same**	**Diff**
Neutral	41	7	5	[1]38
Non-neutral	73	107	61	[2]10
Assortative mating/disruptive/underdominance	33	51	33	2
Overdominance	9	21	7	6
Directional selection: *Gp* → *Gf*	23	20	14	8
Directional selection: *Gf* → *Gp*	8	15	7	4

[1]number of markers that deviated from neutrality in one region but not the other
[2]number of markers that deviated from neutrality in both hybrid zones, but had different patterns of introgression

cline analyses) in both Pennsylvania and Connecticut also have very steep geographic clines in Connecticut (w = 266.4 ± 38.8 m) and are all centered between 286.0 and 338.6 m (solid black lines, Figure 5.4B). Markers that have significantly restricted introgression only in Connecticut have similar cline widths (w = 254.3 ± 44.3 m) and centers (c: 286.0 to 338.6 m). For comparison, we plotted clines of mitochondrial DNA and three anonymous nuclear markers (pUC5, pUC279 and pUC351) from Ross and Harrison (2002). Mitochondrial DNA has substantially more introgression (*G. pennsylvanicus* to *G. firmus*) than the nuclear markers (w = 374.4, c = 588.6 m). One anonymous nuclear marker, pUC279 has asymmetric introgression into *G. firmus* (w = 442.7, c = 399.6 m), while pUC5 (w = 336.6, c = 313.8 m) and pUC351 (w = 316.1, c = 342.5 m) have clines as steep as many of the SNP markers (Figure S5.2).

Discussion

Consistent patterns of introgression throughout the hybrid zone imply strong selection

Stable hybrid zones are products of many generations of recombination between diverged lineages. The success of fragmented and shuffled genomes is dependent on both endogenous and exogenous selection. The strength of selection, as well as linkage relationships, will determine the extent of introgression for a given genomic region; genomic regions that contribute to barriers between species will have low gene flow. Because hybrid zones often extend over large geographic distances, both the nature of the barriers that limit gene exchange and the relative barrier strengths can vary along the zone (Harrison 1990). As a result, both the extent of gene flow for neutral genes and patterns of introgression for individual "barrier genes" can also vary. Genes that contribute to universal barriers should have consistent patterns of

introgression throughout a hybrid zone, whereas genes that vary in patterns of introgression may determine traits that reflect the local ecological context.

We find remarkably consistent patterns of introgression for individual markers between the Pennsylvania and Connecticut samples. Such consistent patterns of introgression between two distant and distinct regions of the hybrid zone must be due to selection. Consistency in pattern occurs in spite of differences in the spatial scale of the samples and in the ecological context. The Pennsylvania samples reflect regional variation over a spatial scale of 10-100 km. The Connecticut samples come a linear roadside transect (about 1 km) across a boundary between loam and sand soils, with which *G. pennsylvanicus* and *G. firmus* affiliate (Harrison 1986; Rand and Harrison 1989; Ross and Harrison 2002). In Pennsylvania, soil is not an obvious determinant of habitat association; *G. pennsylvanicus* crickets are associated with more pristine habitat (forest edges and natural clearings) whereas *G. firmus* is associated with disturbed habitat (agriculture and suburban lawns) (Larson et al. 2013b).

Consistent patterns of introgression have been documented in sunflower hybrid zones (Rieseberg et al. 1999), but in sculpin (Nolte et al. 2009) and mice (Teeter et al. 2010) patterns of introgression vary between transects or samples. The cause of the differences are not clear, although in the case of the house mouse, it has been argued that observed differences could be an artifact of sampling across a patchy hybrid zone and/or population structure (Dufková et al. 2011; Macholan et al. 2011; Janousek et al. 2012). What little variation we see in the field cricket hybrid zone appears to be due to differences between the two samples in the power of genomic cline analyses to detect significant deviations from neutrality. All of the clines with reduced introgression in Pennsylvania have reduced introgression in Connecticut. But 18 clines that do not deviate from neutrality in Pennsylvania have significantly reduced introgression in

Connecticut. These markers have similar distributions of observed genotypes in both transects, with steep genomic and geographic clines. The higher proportion of introgressed individuals in Connecticut increases the power to detect deviations from neutral expectations (Gompert and Buerkle 2009; Payseur 2010).

Gene annotation suggests that some genes with restricted introgression encode proteins with functional roles that may contribute to prezygotic barriers between these species. Six genes encode cytoskeletal proteins associated with actin or tubulin binding, a class of proteins that has been shown to be associated with sperm capacitation and the acrosome reaction during fertilization (Dvorakova et al. 2005; Sun and Schatten 2006). Given the fertilization barrier between *G. firmus* females and *G. pennsylvanicus* males (Larson et al. 2012a; Larson et al. 2012b), proteins that affect sperm function or sperm/egg interactions are candidate "barrier genes." Male seminal fluid proteins that are transferred to females during copulation also likely play a role (Andrés et al. 2006; Andrés et al. 2008; Maroja et al. 2009a). At least one seminal fluid protein, AG-0501F, that was identified through proteomic analysis of the male cricket ejaculate (Andrés et al. 2006) has restricted introgression in Connecticut. Another gene with restricted introgression in both regions encodes a protein with functions common in seminal fluid proteins, but it was not identified in proteomic analyses. For genes that are involved in other prezygotic barriers, such as habitat isolation and assortative mating, it is far more difficult to associate the barrier with a particular protein function. Several genes with restricted introgression encode proteins involved in growth hormone regulation. These genes may contribute to differences in body size and ovipositor length between the two species. Of course, genes with restricted introgression may not be direct targets of selection, but instead may be

linked to gene under selection. Unfortunately, the linkage relationships among the SNP markers and their distribution throughout the genome have not yet been defined.

In crickets, males are the heterogametic sex and are XO. Therefore, males cannot be heterozygous for X-linked SNPs. Thirty-one (27%) of the SNPs we assayed are not heterozygous in any of the males genotyped. The few males identified as multi-generation hybrids are heterozygous at many of the other loci; low overall interspecific heterozygosity is driven by homozygosity for the 31 markers. In addition, all but three males identified as backcrosses have interspecific heterozygosity values below 25%, even when their hybrid indices are proportionally higher or lower (Figure 5.2), while backcross females have interspecific heterozygosity as high as 40%. Of the 31 markers that are consistently homozygous in males, 28 have reduced introgression. Sex linked markers are well documented to have reduced introgression in other hybrid zones (e.g. Payseur and Nachman 2005; Carling et al. 2008; Teeter et al. 2008; Carling and Brumfield 2009; Macholan et al. 2011) and are hypothesized to play an important role in post-zygotic barriers in the heterogametic sex, such as male sterility (i.e. Haldane's rule). However, there is no evidence that hybrids between *G. pennsylvanicus* and *G. firmus* are either sterile or inviable, whereas prezygotic barriers are well documented. The role of sex-linkage in the evolution of prezygotic barriers is less clear (Qvarnstrom and Bailey 2008), but the pattern we see is intriguing and warrants further investigation.

Maintenance of species boundaries in spite of high gene flow

Despite the high levels of introgression in both regions of the hybrid zone, species boundaries remain clearly delineated (Figure 5.3). This is even more remarkable given the scale of sampling in Connecticut. When sampled at a broad scale (10 – 100 km) the hybrid zone

appears clinal, transitioning from *G. pennsylvanicus* in the inland and upland regions to the north and west to *G. firmus* in the lowland and coastal regions to the south and east with only a narrow band of contact along the eastern edge of the Appalachian mountains (Harrison and Arnold 1982). The mosaic structure of the hybrid zone is only revealed at a local scale (1 – 10 km) that corresponds to the scale of patchy environmental variables. Within a few kilometers, transitions occur between pure parental species in adjacent habitat patches. Species interactions occur at these patch boundaries or when migrants move to alternative habitat patches. Therefore, the scale of sampling can have important implications for our estimates of hybridization and introgression. In Pennsylvania, crickets were sampled at random with respect to habitat patches and include individuals both from *G. pennsylvanicus*-like and *G. firmus*-like populations as well as mixed collecting localities (see Chapter 3). At this scale, we find many 'pure' individuals, but we also see evidence of both recent hybridization (one F1 and several multigenerational hybrids) and historical introgression (numerous backcrossed individuals) (Figure 5.2A). In Connecticut, crickets are sampled from across a single patch boundary. At this scale, we see considerably more introgression than in Pennsylvania, yet we find fewer crickets that are products of recent hybridization (no F1s and fewer multi-generation hybrids) (Figure 5.2B). Despite the high levels of gene flow, we find numerous genes with reduced introgression and overall bimodal hybrid indices. Moreover, the geographic transition in allele frequencies is very abrupt (it occurs across a distance of < 100 m), and the majority of geographic clines are centered in an area between 300 – 400 m (Figure 5.4B).

For clines as steep as these to be maintained in spite of gene flow for more than a few generations, there must be strong selection maintaining species boundaries. In both Pennsylvania and Connecticut crickets are associated with different habitat types that maintain

the hybrid zone's mosaic structure. In Pennsylvania there is a patchy distribution of natural habitat (i.e. forest edges and natural clearing) occupied by *G. pennsylvanicus* and disturbed habitat (agriculture and suburban lawns) occupied by *G. firmus*. In Connecticut, habitat patches are differentiated by soil structure; *G. pennsylvanicus* is associated with loamy soils and *G. firmus* is associated with sandy soil. However, habitat is not likely to be the sole factor maintaining the steep clines. Ross and Harrison (2002) compared geographic clines for three anonymous nuclear markers with changes in soil characteristics along the Connecticut transect (% sand and organic content) and found that although the ends of the transect are clearly differentiated as loam or sand, the transition between soil types is much more gradual than the change in allele frequencies. Markers with restricted introgression (based on genomic cline analyses) have similar cline widths and are centered at the same location on the Connecticut transect as two of these anonymous nuclear markers (pUC5 and pUC351) (Figure 5.4B).

These observations suggest that habitat isolation may determine the spatial boundaries between the two species, but that it is a combination of prezygotic barriers (i.e. habitat isolation, mate preference, and fertilization barriers) that restrict gene flow and maintain species boundaries. Clearly, prezygotic barriers between these species are strong. The hybrid zone has a distinctly bimodal distribution (the majority of crickets are either *G. pennsylvanicus*-like or *G. firmus*-like), and there are very few F1 hybrids. We identify 33 markers that have restricted introgression in both hybrid zones, plus an additional 18 markers that have restricted introgression in Connecticut. However, there is extensive gene flow for the rest of the markers (71% in Pennsylvania and 55% in Connecticut). This is a remarkable example of the semipermeability of species boundaries; gene flow can occur throughout some parts of the

genome, but in other regions that contribute to reproductive barriers, gene flow is substantially restricted.

Acknowledgements

We thank E.R. Bondra for assistance with DNA extractions and L. Cote at the Cornell Core Laboratories Center for Sequenom genotyping. This work was supported by a National Science Foundation grant DEB-0639904 to RGH and an AAUW American Fellowship to ELL.

References

Alexander, R. D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). Ann. Entomol. Soc. Am. 50:584-602.

Andrés, J. A., E. L. Larson, S. M. Bogdanowicz, and R. G. Harrison. 2013. Assessing genomic differentiation in a pair of recently diverged field cricket species. Genetics in press.

Andrés, J. A., L. S. Maroja, S. M. Bogdanowicz, W. J. Swanson, and R. G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23:1574-1584.

Andrés, J. A., L. S. Maroja, and R. G. Harrison. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 275:1975-1983.

Apple, J. L., T. Grace, A. Joern, P. S. Amand, and S. M. Wisely. 2010. Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. Mol. Ecol. 19:4012-4028.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113-148.

Bazykin, A. D. 1969. Hypothetical mechanism of speciaton. Evolution 23:685-687.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Ser. B. (Stat. Method.) 57:289-300.

Bridle, J. R., J. Vass-De-Zomba, and R. K. Butlin. 2002. Fine-scale ecological and genetic variation in a *Chorthippus* grasshopper hybrid zone. Ecol. Entomol. 27:499-504.

Carling, M. D., and R. T. Brumfield. 2009. Speciation in *Passerina* buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. Mol. Ecol. 18:834-847.

Carling, M. D., R. T. Brumfield, and M. Webster. 2008. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. Evolution 62:2600-2615.

Dufková, P., M. Macholán, and J. Piálek. 2011. Inference of selection and stochastic effects in the house mouse hybrid zone. Evolution 65:993-1010.

Dvorakova, K., H. D. M. Moore, N. Sebkova, and J. Palecek. 2005. Cytoskeleton localization in the sperm head prior to fertilization. Reproduction 130:61-69.

Fan, S., K. R. Elmer, and A. Meyer. 2012. Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. Philos. Trans. R. Soc. B-Biol. Sci. 367:385-394.

Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. J. Evol. Biol. 23:2004-2016.

Gay, L., P.-A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing cline on molecular and phenotypic traits in hybrid zones: a window on tension zone models. Evolution 62:2789-2806.

Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol. Ecol. 18:1207-1224.

Gompert, Z., and C. A. Buerkle. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Mol. Ecol. Res. 10:378-384.

Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and C. A. Buerkle. 2012. Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. Evolution 66:2167-2181.

Hamilton, J. A., C. Lexer, and S. N. Aitken. 2012. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone. New Phytol. in press.

Harrison, R., and S. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36:535-552.

Harrison, R. G. 1983. Barriers to gene exchange between closely related cricket species . I. laboratory hybridization studies. Evolution 37:245-251.

Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56:337-349.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Pp. 69-128 *in* D. Futuyma, and J. Antonovics, eds. Oxford Surveys in Evolutionary Biology. Oxford University Press, New York, NY.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 111 - 133 *in* D. Otte, and J. A. Endler, eds. Speciation and its Consequences. Sinauer, Sunderland, MA.

Hewitt, G. M. 1988. Hybrid zones - natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158-167.

Janousek, V., L. Y. Wang, K. Luzynski, P. Dufkova, M. M. Vyskocilova, M. W. Nachman, P. Munclinger, M. Macholan, J. Pialek, and P. K. Tucker. 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. Mol. Ecol. 21:3032-3047.

Key, K. H. L. 1968. The concept of stasipatric speciation. Syst. Zool. 17:14-22.

Larson, E. L., J. A. Andrés, S. M. Bogdanowicz, and R. G. Harrison. 2013a. Differential introgression in a mosaic hybrid zone between the field crickets *G. firmus* and *G. pennsylvanicus*. Evolution in prep.

Larson, E. L., J. A. Andrés, and R. G. Harrison. 2012a. Influence of the male ejaculate on post-mating prezygotic barriers in field crickets. PLoS ONE 7:e46202.

Larson, E. L., G. C. Becker, E. R. Bondra, and R. G. Harrison. 2013b. Structure of a mosaic hybrid zone between the field crickets *G. firmus* and *G. pennsylvanicus*. Ecology and Evolution In review.

Larson, E. L., G. L. Hume, J. A. Andrés, and R. G. Harrison. 2012b. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. J. Evol. Biol. 25:174-186.

Lexer, C., C. A. Buerkle, J. A. Joseph, B. Heinze, and M. F. Fay. 2007. Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. Heredity 98:74-84.

Luttikhuizen, P. C., J. Drent, K. Peijnenburg, H. W. van der Veer, and K. Johannesson. 2012. Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica*. Mol. Ecol. 21:3048-3061.

Macholan, M., S. J. E. Baird, P. Dufkova, P. Munclinger, B. V. Bimova, and J. Pialek. 2011. Assessming multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. Evolution 65:1428-1446.

Manel, S., C. Conord, and L. Despres. 2009. Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. BMC Evol. Biol. 9:288.

Maroja, L. S., J. A. Andrés, and R. G. Harrison. 2009a. Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. Evolution 63:2999-3015.

Maroja, L. S., J. A. Andrés, J. R. Walters, and R. G. Harrison. 2009b. Multiple barriers to gene exchange in a field cricket hybrid zone. Biol. J. Linn. Soc. 97:390-402.

Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. Proc. Natl. Acad. Sci. USA 107:9724-9729.

Milne, R. I., and R. J. Abbott. 2008. Reproductive isolation among two interfertile Rhododendron species: low frequency of post-F1 hybrid genotypes in alpine hybrid zones. Mol. Ecol. 17:1108-1121.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. Philos. Trans. R. Soc. B-Biol. Sci. 367:343-353.

Nolte, A. W., Z. Gompert, and C. A. Buerkle. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. Mol. Ecol. 18:2615-2627.

Noor, M. A. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439-444.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316-336.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B-Biol. Sci. 367:332-342.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375-402.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol. Ecol. Res. 10:806-820.

Payseur, B. A., and M. W. Nachman. 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. Biol. J. Linn. Soc. 84:523-534.

Polyakov, A. V., T. A. White, R. M. Jones, P. M. Borodin, and J. B. Searle. 2011. Natural hybridization between extremely divergent chromosomal races of the common shrew (*Sorex araneus*, Soricidae, Soricomorpha): hybrid zone in Siberia. J. Evol. Biol. 24:1393-1402.

Qvarnstrom, A., and R. I. Bailey. 2008. Speciation through evolution of sex-linked genes. Heredity 102:4-15.

Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone - mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432-449.

Renaut, S., A. W. Nolte, and L. Bernatchez. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Mol. Ecol. 19:115-131.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713-727.

Ross, C. L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56:2296-2312.

Schwarz, D., H. M. Robertson, J. L. Feder, K. Varala, M. E. Hudson, G. J. Ragland, D. A. Hahn, and S. H. Berlocher. 2009. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. BMC Genomics 10:633.

Stölting, K. N., R. Nipper, D. Lindtke, C. Caseys, S. Waeber, S. Castiglione, and C. Lexer. 2012. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Mol. Ecol. in press.

Sun, Q.-Y., and H. Schatten. 2006. Regulation of dynamic events by microfilaments during oocyte maturation and fertilization. Reproduction 131:193-205.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution 40:1141-1159.

Szymura, J. M., and N. H. Barton. 1991. The genetic-structure of the hybrid zone between the fire-bellied toads *Bombina-bombina* and *B-variegata* - comparisons between transects and between loci. Evolution 45:237-261.

Team, R. D. C. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien, J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman, and P. K. Tucker. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18:67-76.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker. 2010. The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64:472-485.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Springer, New York.

Vines, T. H., S. C. Kohler, A. Thiel, I. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton, and B. Nurnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B-variegata*. Evolution 57:1876-1888.

Wood, H. M., J. W. Grahame, S. Humphray, J. Rogers, and R. K. Butlin. 2008. Sequence differentiation in regions identified by a genome scan for local adaptation. Mol. Ecol. 17:3123-3135.

Wu, C. I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851-865.

Yanchukov, A., S. Hofman, J. M. Szymura, and S. V. Mezhzherin. 2006. Hybridization of *Bombina bombina* and *B-variegata* (Anura, Discoglossidae) at a sharp ecotone in western Ukraine: Comparisons across transects and over time. Evolution 60:583-600.

APPENDIX

SUPPLEMENTARY INFORMATION

Chapter 3

The following are supplementary tables and figures for Chapter 3.

Table S3.1 Broad geographic population sampling for mtDNA sequencing.

Population	Abr.	Lat. (N)	Long. (W)	Ele. (m)	N
[1] Asheville, NC	ASH	35°33’57”	82°29’49”	640	3
[1] Bishopville, SC	BSV	34°11’29”	80°10’05”	60	2
[1] Blue Ridge Parkway, VA	BRP				3
[3] Cape May, NJ	MAY	38°56’39”	74°51’50”	2	10
[2] Cornwall, PA	COR	40°16’05”	76°24’16”	183	3
[2] Covington, VA	COV	38°00’50”	78°28’21”	354	6
[2] Durham, NC	DUR	36°03’23”	79°04’45”	159	2
[3] Eastham, MA	EST	41°51’58”	69°59’24”	13	10
[3] Elk Neck, MD	ELK	39°28’57”	75°59’13”	32	10
[2] Essex, MD	ESS	39°18’20”	76°28’46”	0	5
[3] Fort Mott, NJ	MOT	39°37’10”	75°33’39”	13	10
[2] Franklin, WV	FRN	38°39’20”	79°19’59”	551	4
[1-3]Guilford, CT	GUI	41°16’48”	72°42’02”	0	21
[1-3]Ithaca, NY	ITH	42°26’01”	76°29’59”	250	21
[1] Lexington, SC	LEX	33°57’25”	81°13’51”	124	1
[2,3]Moorefield, WV	MOO	39°04’09”	78°55’58”	285	10
[2] New Bloomfield, PA	NBL	40°28’24”	77°11’33”	0	6
[1] Pantego, NC	PAN	35°37’07”	76°38’46”	2	3
[2,3]Parksley, VA	PAR	37°45’58”	75°36’00”	0	13
[3] Point Judith, RI	PTJ	41°21’40”	71°28’53”	0	19
[3] Pownal, VT	POW	42°45’35”	73°13’59”	169	10
[2] Ritchie, MD	RIT	38°52’07”	76°51’01”	0	6
[2,3]Scranton, PA	SCR	41°24’25”	75°35’46”	397	12
[1] Sharon, CT	SHA	41°52’45”	73°28’36”	216	4
[2] South Hill, VA	SOH	36°45’07”	78°06’09”	116	4
[2,3]State College, PA	SCO	40°47’59”	77°52’05”	371	12
[2,3]Tom’s River, NJ	TRI	39°45’00”	74°11’33”	0	15
[1] Wrightsville, PA	WTV	40°01’39”	76°33’09”	120	3
				Total	**228**

[1] Willet et al. (1997); [2] Maroja et al. (2009a); [3] this study

Table S3.2 Amplicon primer sequences, expected amplicon product length (bp), annealing temperature (T°C), extend primer sequence and target SNP for each mtDNA SNP assay.

SNP	Amplicon sequence (5′ - 3′)	bp	T°C	Extend sequence (5′ - 3′)	SNP
70[2]	F:CCTCCTAGGAAGATTAGGAA R:CAGGATCAATCTCTTACTC	81	48.8	R:CCTAGGAAGATTAGGAATAGGATATA	T/C
310[2]	F:TACATTGATTAAGAAGGAAG R:GCATTGATTATAATTCTGTG	101	45.7	R:tttTTAGTCGTAAATAGTAGAATAAAGT	T/C
448[2]	F:CCAAGTGAGCAAATATATG R:GCCATTTAACATTTACCTT	119	46.3	R:agTATGAATGTTAATCAGAATGGTAA	G/A
454[2]	F:GGTGTAATCCCTAAAGGTGG R:TCATTTCATTCTTTTGAGC	108	46.5	R:GATATTAGGAGCTAAACTTCTATG	T/C
665[2]	F:GATTTTGACTTCTACCCCCG R:TGTTCATCCTGTTCCTGCAC	100	46.8	F:CCTTTTATTAACCAGAAGAATAGT	C/T
796[1]	F:TGGAGGTTTAACAGGTATTG R:ACGTAATGAAAATGGGCAAC	110	45.3	F:TAACAGGTATTGTTCTTGC	C/T
952[1]	F:CCTTTATTCACAGGATTAAC R:GAAAGTGTTGTGGAAAGAATG	120	45.6	F:GATTAACAATAAATCCTAAATGATTAAA	A/G
1036[1]	F:CATGATGTGTAAGCGTCTGG R:CACTTTCTTGGATTAGCAGG	85	46.3	R:CGTCTGGATAATCGGAATA	G/A
1084[2]	F:ACCAGCTAAAACTGGTAGTG R:ACGAGCACCAGGAATATCAC	120	46.5	F:cTGATAATAATAATAGAAGAGCTGTAAT	C/T
1204[1]	F:GAATAGGAGTGTTCTGCTGG R:TCGAAAGCTTATATTTCCC	101	52.1	R:ctcGTTCTGCTGGAGGAAGATTTTG	T/C
1382[1]	F:TACAAAACAGTTCATCACCC R:AATATATGAAACGAGAATTG	118	46.9	F:TTCATCACCCTTAATAGAACA	A/G

Table S3.3 Samples sizes and mean morphological measurements for allopatric populations of *G. firmus* and *G. pennsylvanicus*

		Males							**Females**				
	Pop	N	BL	PW	FL	B	C	H	N	BL	PW	FL	OL
Gf	GUI	20	21.3	5.6	11.6	133.9	29.3	0.2	13	21.7	5.8	12.5	20.3
	TOM	27	20.7	5.9	12.0	142.4	36.7	0.5	21	25.0	6.4	13.7	20.7
	MAY	3	19.4	5.3	12.0	133.8	42.2	0.6	8	21.6	5.8	12.7	18.9
	MOT	12	21.5	6.3	13.4	159.2	37.8	0.4	5	26.0	6.1	14.0	20.8
	MET	16	21.6	6.3	12.9	160.1	33.4	0.2	13	23.3	6.6	13.8	19.5
Gp	ITH	37	18.9	5.4	10.9	92.2	18.1	0.2	8	18.5	5.1	10.7	14.0
	NBL	20	21.5	6.3	12.0	102.0	14.5	0.2	10	27.1	6.8	12.2	16.1
	SCR	6	18.2	5.1	10.7	72.4	10.4	0.2	2	22.7	5.6	12.7	14.6
	SCO	9	19.8	5.6	11.6	74.1	15.4	0.3	6	24.4	5.8	12.7	14.6
		182.9							210.3				

Table S3.4. Mean morphological measurements for hybrid zone populations.

	Males						Females			
Pop	**BL**	**FL**	**PW**	**B1**	**S5c**	**H4c**	**BL**	**FL**	**PW**	**OL**
A	22.1	12.0	6.3	46.5	8.8	0.3	20.4	11.1	5.5	14.9
B	22.4	12.2	6.4	43.8	13.5	0.5	17.6	11.1	5.5	16.5
C	19.5	10.5	5.2	69.6	14.2	0.5	19.4	11.5	5.5	14.3
D	19.5	11.3	5.4	96.5	20.5	0.5	19.9	11.7	5.4	15.1
E	20.8	12.0	5.9	91.7	18.6	0.5	21.6	10.9	5.7	14.8
F	21.4	11.3	5.9	71.7	12.8	0.3	21.9	12.2	5.9	15.7
G	21.2	12.2	6.0	98.2	24.9	0.5	20.6	11.9	5.8	16.3
H	20.6	11.5	5.6	71.2	11.0	0.2	21.9	11.5	5.5	14.9
I	23.0	12.9	6.6	133.7	34.2	0.5	24.1	13.4	6.5	19.0
J	21.2	11.3	5.7	102.8	29.4	0.5	20.6	12.2	5.6	15.6
K	23.4	13.5	6.6	150.1	40.5	0.5	23.6	14.3	6.8	21.0
L	20.3	11.2	5.5	113.4	25.8	0.4	20.9	12.7	5.8	16.2
M	18.2	10.3	4.9	106.4	20.9	0.4	19.2	11.0	5.2	14.8
N	21.6	12.4	6.4	75.6	15.1	0.2	22.0	12.0	5.9	16.1
O	21.5	11.8	5.9	89.4	19.7	0.2	-	-	-	-
P	22.8	11.7	6.0	85.6	12.0	0.0	23.8	12.4	6.1	16.0
Q	-	-	-	-	-	-	21.1	12.1	5.6	16.8
R	20.6	11.4	5.4	86.2	16.8	0.2	20.8	11.3	5.4	14.9
S	-	-	-	-	-	-	26.2	13.2	6.7	16.8
T	20.2	11.9	5.7	163.8	44.2	0.7	-	-	-	-
U	21.4	11.5	5.6	123.0	35.4	0.5	24.7	12.8	5.9	18.1
V	23.4	13.4	6.4	129.1	39.3	0.5	-	-	-	-
W	21.8	13.2	6.6	136.5	36.2	0.6	-	-	-	-
X	23.9	12.8	6.7	100.1	28.7	0.2	26.0	13.5	6.5	18.3
Y	23.4	13.1	6.6	174.1	43.4	0.6	25.8	13.5	6.7	18.1
Z	21.9	12.3	6.2	157.8	40.8	0.6	-	-	-	-
AA	20.1	11.5	5.7	89.0	20.5	0.4	-	-	-	-
AB	21.1	11.3	5.5	106.1	24.0	0.5	-	-	-	-
AC	20.9	12.5	6.0	129.5	30.2	0.5	21.7	11.9	5.9	14.7
AD	19.6	12.1	5.9	130.5	36.4	0.5	24.2	12.0	5.7	17.4
AE	21.2	11.6	5.8	84.1	15.5	0.2	24.2	12.9	6.4	18.3
AF	21.1	12.7	6.2	99.8	22.5	0.2	22.3	12.7	6.3	15.5
AG	21.8	12.5	6.2	164.9	35.5	0.2	20.2	11.3	5.4	15.6
AH	18.4	10.1	4.7	73.2	13.2	0.2	-	-	-	-
AI	20.3	12.8	5.6	139.0	32.5	0.4	22.9	12.8	6.3	17.1
AJ	19.0	10.5	5.0	115.8	22.9	0.0	19.8	12.3	5.6	14.8
AK	18.7	11.1	5.1	107.8	19.7	0.1	21.8	11.5	5.4	13.8
AL	22.8	13.7	6.7	115.9	30.1	0.2	25.2	14.6	6.7	19.7
AM	20.1	12.0	5.8	174.3	35.7	0.3	23.1	13.3	6.4	18.5

Table S3.4 (Continued)

AN	17.1	9.9	4.6	110.9	17.9	0.2	20.5	11.1	5.7	14.0
AO	19.9	11.9	5.7	158.4	37.8	0.4	25.1	14.2	6.9	20.5
AP	21.2	11.7	5.6	202.0	45.2	0.6	25.8	15.6	7.3	20.3
AQ	22.2	12.6	6.4	145.4	35.4	0.2	25.0	13.5	6.5	18.1
AR	21.2	12.5	6.1	121.0	23.3	0.3	22.7	11.2	5.4	12.8
AS	22.3	12.7	6.3	180.2	45.1	0.5	24.4	13.7	6.6	19.1
AT	-	-	-	-	-	-	23.7	14.1	6.2	18.4
AU	21.1	13.0	6.0	121.4	31.6	0.4	22.4	13.2	6.1	18.5
AV	21.6	13.1	6.2	136.5	44.2	0.6	26.5	14.6	6.7	20.2
AW	22.5	13.4	6.3	161.5	34.4	0.4	25.0	14.3	6.7	19.6
AX	19.9	12.3	5.6	89.4	20.4	0.3	22.5	12.3	5.8	14.6
AY	22.4	13.1	6.4	165.7	35.8	0.4	27.0	14.1	6.6	20.3
AZ	23.0	13.7	6.6	172.0	34.2	0.1	24.4	14.3	6.6	19.0
BA	22.8	13.9	6.6	143.7	42.6	0.7	26.0	14.3	6.3	19.5
BB	21.7	13.0	6.1	137.4	43.4	0.7	25.8	13.3	6.1	18.3
BC	22.9	13.2	6.3	133.6	42.0	0.7	25.4	13.9	6.5	19.2
BD	23.5	14.1	6.8	115.4	38.9	0.6	23.7	14.3	6.4	20.3
BE	22.5	13.3	6.3	124.4	27.2	0.4	-	-	-	-
BF	19.4	13.8	6.0	84.7	19.4	0.4	24.1	13.9	6.2	15.7
BG	24.4	13.3	6.4	111.1	32.7	0.5	25.4	14.1	6.4	19.9
BH	19.4	11.9	5.5	118.6	37.8	0.6	21.6	12.1	5.6	16.8
BI	20.5	12.4	5.9	87.5	22.0	0.3	23.0	12.7	5.9	15.5
BJ	22.7	11.5	5.1	101.6	30.0	0.6	22.2	12.1	5.7	14.9
BK	19.8	11.7	6.2	89.8	25.4	0.5	22.7	13.5	6.2	16.9
BL	22.5	13.5	6.6	143.0	45.0	0.6	23.9	13.9	6.4	19.9
BM	22.0	13.6	6.5	142.0	41.5	0.7	24.0	14.2	6.5	20.4
BN	20.7	12.2	5.8	131.3	34.5	0.5	24.1	13.2	6.0	19.0
BO	23.6	14.3	6.8	167.8	36.8	0.4	24.8	14.2	6.6	19.2
BP	20.0	11.9	5.6	153.9	36.3	0.4	21.9	13.0	5.9	17.6
BQ	21.7	12.6	6.1	164.8	42.7	0.5	24.8	13.7	6.3	19.7
BR	23.2	12.8	6.4	102.4	36.6	0.5	26.0	14.2	6.6	20.1
BS	20.9	12.8	5.8	123.2	46.2	0.6	26.1	14.5	7.1	19.9
BT	22.7	13.4	6.7	152.2	43.7	0.6	27.0	14.7	7.0	19.9
BU	21.3	12.6	6.1	142.3	33.6	0.6	23.0	12.1	5.9	15.7
BV	22.0	13.5	6.3	135.1	34.1	0.3	25.3	13.9	6.5	19.5
BW	23.3	13.9	6.8	123.2	32.6	0.4	26.0	14.2	6.6	20.4
BX	22.0	13.3	6.3	144.0	31.5	0.3	24.0	13.5	6.4	18.9
BY	20.2	12.8	5.8	90.7	18.2	0.4	23.7	13.0	5.7	14.5
BZ	20.2	12.8	6.0	151.6	27.9	0.3	24.7	12.8	6.0	16.2
CA	19.9	11.5	5.6	103.2	23.5	0.6	24.6	13.2	6.2	16.6
CB	20.5	12.1	5.9	82.1	17.2	0.4	24.1	12.1	5.8	14.5

Table S3.4 (Continued)

CC	25.0	14.1	6.9	79.4	33.5	0.5	27.7	15.5	7.0	21.5
CD	22.3	12.5	6.3	89.0	23.6	0.5	26.8	13.8	6.4	18.7
CE	21.6	13.0	6.3	103.8	25.6	0.6	26.7	14.7	6.9	18.8
CF	21.1	12.8	6.0	65.1	11.4	0.4	26.0	13.3	6.5	15.7
CG	24.1	13.9	6.7	128.8	35.4	0.6	24.3	14.6	6.8	20.1
CH	22.0	13.1	6.6	85.2	22.7	0.3	25.7	12.8	6.4	17.4
CI	21.1	12.7	6.1	114.7	23.6	0.2	24.4	13.7	6.4	17.0
CJ	21.0	11.6	5.8	101.6	18.0	0.0	24.7	12.0	5.9	15.8

Table S3.5 Results of simple linear regression for environmental variables influencing ovipositor length and morphological cluster membership in populations of *G. firmus* and *G. pennsylvanicus* in Pennsylvania.

Ovipositor length	**$F_{[1,77]}$**	**R^2**	**P**
Natural vegetation	26.795	0.258	<0.0001
Latitude	16.961	0.180	<0.0001
Vegetation density	13.054	0.145	0.0005
Annual Temperature	11.570	0.130	0.0011
Annual rainfall	6.915	0.082	0.0103
Sand	4.110	0.050	0.046
Elevation	2.139	0.027	0.147
Human footprint	0.027	0.003	0.600
Topographic complexity	0.048	0.001	0.826
Silt	2.367	0.029	0.128
Clay	0.022	0.0003	0.654
Organic content	1.766	0.022	0.187
Membership Coefficient	**$F_{[1,86]}$**	**R^2**	**P**
Natural vegetation	18.912	0.180	<0.0001
Latitude	31.165	0.266	<0.0001
Vegetation density	5.004	0.055	0.028
Annual Temperature	9.473	0.099	0.0028
Annual rainfall	5.998	0.065	0.016
Sand	4.529	0.050	0.036
Elevation	0.312	0.003	0.577
Human footprint	0.011	0.0001	0.916
Topographic complexity	0.366	0.004	0.547
Silt	4.335	0.048	0.040
Clay	1.132	0.013	0.290
Organic content	1.736	0.019	0.191

Table S3.6 Model selection for environmental variables influencing ovipositor length and morphological cluster membership in populations of *G. firmus* and *G. pennsylvanicus* in Pennsylvania.

Model	R^2	AICc
Ovipositor length		
Lat, VegDens, NatVeg, Lat * NatVeg, VegDens * NatVeg	0.5632	294.365
Lat, VegDens, Org, NatVeg, Lat * NatVeg	0.553	296.199
Lat, VegDens, Org, NatVeg, VegDens * NatVeg	0.5526	296.258
Lat, VegDens, Clay, NatVeg, VegDens * NatVeg	0.5508	296.570
Lat, VegDens, NatVeg, Lat * VegDens, VegDens * NatVeg	0.5492	296.861
Morphological cluster membership		
Lat,Org,NatVeg,Lat*NatVeg,Org*NatVeg	0.5244	13.105
Lat,sand,Clay,NatVeg,Lat*NatVeg	0.5210	13.744
Lat,Silt,Clay,NatVeg,Lat*NatVeg	0.5197	13.980
Lat,Clay,NatVeg,Lat*NatVeg,Clay*NatVeg	0.5196	14.001
Lat,VegDen,Clay,NatVeg,Lat*NatVeg	0.5185	14.193

We compared all possible models using Akaike Information Criterion (AICc); five best models are reported for each dataset. Best predictors are: Latitude (Lat), Vegetation Density (VegDens), Natural Vegetation (NatVeg), Organic Matter (Org), Clay (Clay), Sand (Sand), and Silt (Silt).

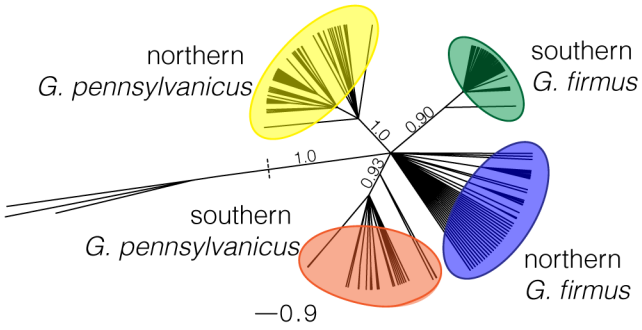


Figure S3.1 Bayesian 50% majority-rule consensus tree of the mtDNA gene cytochrome oxidase I sequence data. Values on the branches correspond to the Bayesian posterior probabilities. The tree includes all samples from Willet et al. (1997) and Maroja et al. (2009a), plus an additional 130 crickets from 13 localities across the hybrid zone (Table S3.1) and 119 crickets from 31 localities within the Pennsylvania hybrid zone (Table 3.1).

.

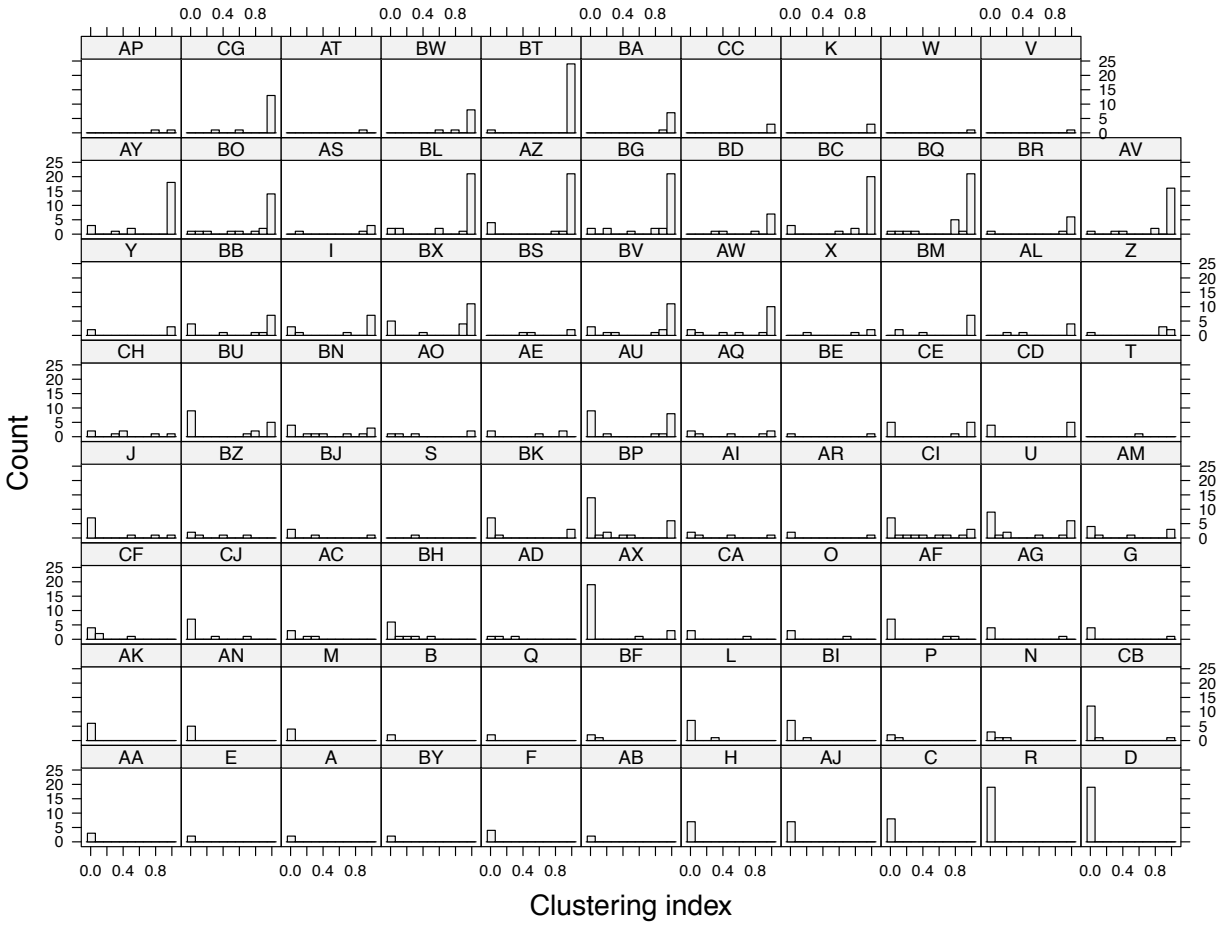


Figure S3.2 Distribution of morphological clustering membership coefficients for each collecting locality. Indices are based on fuzzy c-means clustering of morphological measurements for males (body size and tegmina color, $r = 1.25$) and females (body size and ovipositor length, $r = 1.5$). Collecting localities are ranked by the average clustering index, with *G. firmus* localities at the top-left and *G. pennsylvanicus* localities at the bottom-right

Chapter 4

The following are supplementary tables and figures for Chapter 4.

Table S4.1 Sampled allopatric and mixed populations of *Gryllus pennsylvanicus* and *G. firmus* used for SNP validation. Crickets from these populations were initially described in Maroja et al. (2009) and Larson et al. (2013).

Population	Abr.	Latitude (N)	Longitude (W)	Elevation (m)	N
G. pennsylvanicus					
Ithaca, NY	ITH	42°26'01"	76°29'59"	250	12
Scranton, PA	SCR	41°24'25"	75°35'46"	397	11
State College, PA	SCO	40°47'59"	77°52'05"	371	12
G. firmus					
Guilford, CT	GUI	41°16'48"	72°42'02"	0	12
Tom's River, NJ	TRI	39°45'00"	74°11'33"	0	12
Parksley, VA	PAR	37°45'58"	75°36'00"	0	12
Mixed					
Moorefield, WV	MOO	39°04'09"	78°55'58"	285	9

Table S4.2 Sampling locations and number of crickets collected per locality (N) for the southern Pennsylvania region of the hybrid zone. Crickets and collecting sites were initially described in Larson et al. (2013).

Locale	N	Latitude (°N)	Longitude (°W)	Elevation (m)
AE	5	40°24'17"	76°29'51"	244
AJ	5	40°30'36"	77°17'42"	359
AK	5	39°58'80"	77°22'46"	496
AQ	5	39°49'18"	77°57'12"	190
AW	5	39°54'49"	77°54'19"	209
AX	9	39°55'45"	77°57'18"	696
AZ	6	39°50'36"	77°28'40"	510
BA	5	39°48'38"	77°34'80"	235
BG	5	39°56'51"	77°33'41"	258
BK	5	39°44'22"	77°28'11"	442
BN	5	39°52'70"	77°38'90"	250
BO	5	39°52'21"	77°32'36"	272
BP	5	39°50'40"	77°42’10"	195
BQ	5	39°52'20"	77°43'41"	201
BU	9	39°51'23"	77°51'41"	174
BV	10	39°51'12"	77°48'40"	163
BX	5	39°57'34"	77°47'54"	245
BZ	5	39°59'16"	78°30'45"	250
CA	4	40°50'16"	78°31'14"	355
CB	5	40°10'25"	78°31'42"	354
CD	9	39°59'50"	78°60'37"	322
CE	5	39°55'43"	78°60'28"	290
CF	6	39°49'40"	78°15'15"	286
CG	5	39°59'57"	78°14'14"	406
CH	6	40°00'31"	78°22'35"	345
CI	10	40°80'29"	78°19'59"	297
CJ	5	40°20'70"	78°22'50"	397
D	5	40°15'51"	76°26'80"	220
H	5	40°26'57"	76°13'58"	140
I	12	40°22'51"	76°10'46"	105
J	5	39°44'15"	76°20'48"	102
L	5	39°48'49"	76°19'42"	149
N	5	40°32'70"	76°57'59"	204
R	5	40°53'2"	76°48'16"	204
U	5	40°30'49"	76°42'34"	141
Y	5	40°40'60"	76°54'32"	176

Table S4.3 Amplicon primer sequences, expected amplicon product length (bp), annealing temperature in degrees Celsius (T), extend primer sequence and target SNP for each contig in nine MassARRAY iPLEX assays designed for *G. pennsylvanicus* and *G. firmus*.

	Amplicon primers (5′ - 3′)			Extend primer (5′ - 3′)	
Contig	Sequence	bp	T	Sequence	SNP
3344	F:TTTTGTCGGCCGAATACGTG R:GAGTAGTCTTTTGCGGTGTC	118	48	R:CTTGCACTCGCGTTT	C/A
2831[2]	F:ACTATTACGGCTAATCGCGG R:GCCAACCTATATTAGCCTAC	175	46	R:CGCGGGAGAAGTAAG	G/A
2658	F:ATCACGGGAAACCTATTGGG R:GGTTTTGTCCGAAGAAGAGG	116	48	F:tCAAGCACTCTCCTCCA	C/A
5961	F:TCATCTCCTCACCGTCATAG R:ATGCAGACTGCAACAGTCTC	136	52	R:taATGGCCCCTCTGCAC	G/A
3432[2]	F:CGACAGAAGTCAAAACTGAG R:CACCATTCCACGTTACAAGC	117	46	R:ACTGAGTAGCCTCACTA	G/A
5214[3]	F:CGGAAAGATCGGACGTGAAT R:AGACTTATGCCGTAGCACAC	130	45	F:agCTGGATGTCGGACTG	C/T
1851	F:GTTGTCGCCTGTTTCTATAC R:TGCTCCTACATTTGCAACAG	128	46	F:CACTTGTTGGGTTAGGTA	C/T
4205	F:ATCACAGAGAAAGACGCAAG R:GGGAGTGATTGCTGAATGTG	128	48	R:GCTAATTGCATCTCCACAA	G/A
5214[1]	F:ACTTCACACAACTCGAGCAG R:TGCTTGCAGTTATGCGTTGG	135	47	R:ccGTATGCTGGGACAAGTT	T/A
14741[1]	F:CTACTGTTCCTATTGGCACC R:GGATTTGCAAATAGTCCTTC	156	50	F:ctaCTATTGGCACCCCACAG	C/A
14741[2]	F:CTACTGTTCCTATTGGCACC R:GGATTTGCAAATAGTCCTTC	156	45	F:ctgGTGTGGCTGAAAATGAT	C/A
2831[1]	F:GCCAACCTATATTAGCCTAC R:ACTATTACGGCTAATCGCGG	175	45	F:CCTATATTAGCCTACAAGTTC	C/T
8375	F:ATGGAGCTTCACCATGATGT R:AACCATTGTTGAAGTGAGGC	112	47	R:AATTCTAAAGCTGCAAATCTC	G/A
402	F:AAAGAAGGAAAGGGTGCGTC R:GCATTGTTGCATAGTCTGCG	137	47	F:cGGGATGAGAACAAGTGAATA	C/T
1231[3]	F:GCATCAATGCTGTAATGAAG R:GGAAAAATATCCTACTCTGCG	164	47	R:agACTGCTCAAATTTAATTCGG	T/C
5368[1]	F:GCCGTTAGCTTTGAGTCTTC R:CTGGTGTTCCTACACGTTTC	120	49	F:ttATGGTGATACGGAAATTCCT	G/T
5214[2]	F:ATGGTTGAAATCAGCCTTGC R:GATCAAGGAAGTGATGTCAG	164	47	F:TGTATTGCTGAGGAAAATTAAG	G/T
7566	F:GTTACGTGTGACGAAGTGAC R:CACTCAGAATGCACAAGAAG	121	49	F:cacTAGAACTCTTGGCTCTTCAA	A/T
1412	F:TCGTTTCTCCAACTTGCCAG R:CTTTTAGATAAATGGCCAGG	138	46	R:gAAGAATCTATATATTCCCAGGTA	G/A
726	F:CCAGAAGACAAACCTTCAAG R:AAACTCGCCTCCTATCAGTC	153	49	R:CGGATAATAGGATGAGAAACATAG	G/A
1539	F:GTCTGTTTTAACAAGTGGTGG R:TGCAATATGTGAGAACTGGG	132	48	F:cggTAACAAGTGGTGGAAAGTAAT	A/G
5131[1]	F:CAACAATTGGCGACTGAAGG R:CTGGTTTACTGAAGTCAGCG	127	48	R:TTGACTATTACCATGACAGTATTAA	G/A
1231[2]	F:GCATCAATGCTGTAATGAAG R:GGAAAAATATCCTACTCTGCG	164	55	R:cccACCCATTCCTTGATTCCAGGTAG	G/A
9839	F:TGAAGAAGATATCACAGCAC R:TCTTCCTAACCGAAACACAG	121	50	F:cTATTTCTTAGTCTCCTTTTCTTGGT	A/G
7153[1]	F:CTACAGCTGGAGATATTGGC R:GTTACTGGCTCCTGCATTTC	123	54	F:tAGATATTGGCTCAGCAAATGTCAAA	C/A
7046[1]	F:GACTTCCGGTGTTCAGAAAC R:TCGTCACCTGATAAAGGAAC	140	48	F:AACAAAAAGAAAAATGATAAGAAACC	G/T
3104[2]	F:TTCCTGTCAAGTAGATTCGG R:TGTTCTTTAACGGTTGGAGC	110	46	R:tcATTTTGTCAATATGTTTAACCTATT	G/A

Table S4.3 (Continued)

5067	F:TGGAATCTTCAGCTGACTTC R:ATGTGTTGCCAACTACGACC	134	52	F:ttATCTTCAGCTGACTTCTTATGTTTG	A/G
1374[1]	F:TTGCTTAGAGGCTTTGCTCG R:GCCGGAGACATGATGAGAAA	142	52	R:gagTTTGTTGGGGGCAATAAAATACAT	G/A
1513	F:CTCGCTTTCTTTCGAATCCC R:GGAACCGTGGACTTGAATAC	126	48	R:cccACGACATTGTTTGTAAGTATTAAAA	G/A
7046[2]	F:TCAACGACGATAGTTTCAGG R:TCCTTTTTCTGCCAAACCCC	128	49	R:gcaGTACTTCACAGATTTCAATAGTTTG	T/C
1234	F:CGGGATAGTTGAGAAAGTTG R:AATGTGTCACAACCTCAACC	126	48	F:AGAAAGTTGTATGAATTGAATAATTAAC	C/A
1374[2]	F:GCCGGAGACATGATGAGAAA R:TTGCTTAGAGGCTTTGCTCG	142	48	F:CCCCCAACAAAAGCC	A/T
2864[2]	F:CGTCAAAAGCCAAAGAGAAC R:TTTTTTCAGGCGGCATGGAC	137	49	R:ACCTGGGTGGTTTGG	T/A
3555	F:GCCCGTATTTCTCCTTCATC R:AGGGCCAATTTCGACTTCTG	123	46	R:CAGCAGCAAAACTGTC	G/A
10368	F:TAAAATCTCAGCGCCAACCG R:GCTCAAACTTGGAAGGAGAC	120	51	R:cTTGACGGAAGCCCCAT	C/A
2864[1]	F:TTTTTTCAGGCGGCATGGAC R:CGTCAAAAGCCAAAGAGAAC	137	46	F:GACCATGATAGGGGTTT	C/T
5131[1]	F:TTGGAGTGATTCGTTCAGGG R:GTCAGCCACTCAAAATACTG	120	47	R:aGTTCAGGGCAAGTTGA	T/C
5177	F:TTGTCAGTCTTCAGACTGGG R:GAAGAAGGAGATGGAGACAC	121	47	R:TGGGAGGAAATAGGTGA	T/A
1032[1]	F:TCATTTGAGCCATTAGCTCC R:GCATATTCACTTCCTGCCAC	120	50	R:GCTCCTCTAGGGAAACATC	G/A
425	F:GAAACTCGAGAAGTCCTTCC R:ACTCGGGCTCACTATTAGTC	120	46	F:GGATAAACAGATCCAGTCT	C/T
7153[2]	F:CTACAGCTGGAGATATTGGC R:GTTACTGGCTCCTGCATTTC	123	45	F:gTAAGGAAGAAGAAGCTGT	A/T
8257	F:CTACGTTTCAGTGGCAAATC R:AATCTTCTCTTGCTTCGTGC	151	46	R:AGTGGCAAATCATAGTAATG	T/A
7469	F:AGTGCAGGAAGTAATGTGGC R:TGAAAGGCACCAATCCAAGG	120	45	F:ctATATGCCAATTGTGTATCC	C/A
94	F:ATTTATGCAAGCTTCTCTGG R:GGAGCAAATGATCTGCTAGG	119	47	R:AATGCACAAGAACATTATTCA	T/A
AG-0383[2]	F:GCCAATTCATATTTTTGAG R:CCGTGGACACTGCTAACTTA	120	47	R:gTTTTTGAGAAGTAGGAGCAA	T/A
11695	F:CGTAGGGTATGGAGAAAAGC R:AAAATGCAAGGCTGGGCAGG	139	49	R:tAAGGATGCCAATGGTTAATAC	G/T
6571	F:CAAGTACAGACAACAATGAC R:TCTGCACCAATCAGTCAATC	148	46	R:cccATGTTCTAGCATTTACAACG	G/A
618[1]	F:ATTGCCCAAGGACGCTGTTC R:GGCGGATATGCAGATGATAC	136	45	F:ATGATTTAATTTCTTCAGTTTCA	C/T
16015	F:CTCGATCTAGAGCTGATGAC R:TTAGTGATGATGAGGCACCC	161	49	F:cacCACAATCTCATCATCTTCGTC	A/G
5556[1]	F:AAAAGCAAAAAGGGTGGGCG R:CATCACTTTTGGCTTTCTTC	141	56	R:ttAGCAGGCTGCTAAAAAGCCTAA	T/C
2733	F:CATTAAGTTGTCTTTACACCC R:CTCTTACACGTAGGTGATTG	125	50	F:cTTCTTGTAATCTAACTGCTCTCTT	A/T
6030	F:ACGCCCCCCTTTTGAAATTC R:AAAGATTCCCTGGACCGAAG	133	51	R:TTCAAGGCAATAGGAAGAAAATACT	G/A
4481	F:ATCATCGGTTCTATTGGTGG R:GAGTTGTCATAAGTCCTCTG	148	48	R:CGAAAAATTATAACCTATTCATCCAT	G/A
1369[1]	F:GGGCCATACTCTTTGCAAAT R:CATTTTAAAATGACCACCAG	163	50	R:TGCAAATATCATCAGTGATATTATCC	G/A
827	F:CCTTGCACATGCCATTCATC R:GTATACACAACTGAAATGC	103	46	F:ggAATAACTGATCAAAACAGTTAATT	G/T

Table S4.3 (Continued)

1555[1]	F:GTTCACGTTCTCTGTCTTGG R:AGGGCCAATTTCTACCAGAG	108	52	F:ggGAATTTGGTTCCGGAATTATCAGAT	A/T
3313	F:CGCACAGTCATGATACATTC R:CAAGCGAAGTGTGAAGTGTG	146	51	R:cTTCAGTCGTAAGAACCTATTCTTTATT	G/A
17333[1]	F:TGTTTTAAGACATTCCTGCC R:GCATTCTTGTATTTCTTTC	148	47	R:aacTTTGTATCTAATAGCAATCCTTTAT	T/C
12397	F:CAGCAGCTATATAGTAACCTG R:TCCACCAACCCACATGTACC	126	49	R:ggAACCTGATAACCTTATAGGAATAATC	T/C
874[2]	F:AACTAGGTGTCACCGGTTTC R:GTCAGGTTGGTTTTCTTCCG	108	53	F:CCGGTTTCGCGCACT	G/T
11867	F:CACAACGACCTGTTGTAATG R:CCTGACTTTTCTTTTGGTCC	132	48	F:gAGGTGCTCCTGATGC	C/A
5556[2]	F:CATCACTTTTGGCTTTCTTC R:AAAAGCAAAAAGGGTGGGCG	141	48	F:ccAGCAGCCTGCTTCTT	C/G
8612	F:TTTCCGTCCTTCTCACTTCC R:ATTGGTTCCGAAGACAGATG	122	45	F:CTTCACACCTTTTGGATT	C/A
2989	F:AAATCTCGTTTCCAGTAAG R:GAAGACACGTTTGCTCTCAC	136	46	F:CCTATTTTTACGTTGGGC	C/A
2182[1]	F:TGTTCTGCAACATGATTGTG R:GACCAAGCAAGGGAAAATCG	130	47	R:AACATGATTGTGGTGACA	G/A
8354	F:ATGCACACTTTCCATCCAGC R:TTGATGTTCGGACCGAGTTC	148	47	F:gAGCGTATAAGCATGACAC	A/T
1231[1]	F:GGAAAAATATCCTACTCTGCG R:GCATCAATGCTGTAATGAAG	164	49	F:ATCCTACTCTGCGAAGTTTA	C/T
5368[4]	F:CCCTGAAAAAGAGGAAAAGC R:AGCACTGGATTCTGAACTAC	156	49	F:AGAGCACTTGATAATGTGGA	A/G
3422	F:TGTGCCTGCATGCATGATAC R:ACATCAACTCTACCATGACC	119	49	R:aaGGGGGATCAGGATTGAAA	G/A
963	F:ATAGGTGGGTTACGTGAGAC R:TCACTTGCGACTTTTGAGAC	130	51	F:TGGCATTTGATTTGCAACTTG	C/T
2361	F:GGGCTTTTAGCCGAAACTTG R:TGAGGTAAACTATCACCCCG	122	52	F:TAGCCGAAACTTGCAGTAAAAC	G/T
AG-0501F[1]	F:GCCTGATACAGCATTGCTTG R:CTCTGACTTCCATTTCTCCC	133	48	F:gAAATGCAGCAAAGTTAAAAGA	A/G
211	F:GTAACAGCCTCATTACTTCG R:GGGCAAACAAAGTAAAAGAAC	104	47	F:CTCATTACTTCGAAGGATATTTC	A/G
4361	F:ACGAAGCATCGCTGTAAAAC R:TAGGAAGACCACGATACACC	122	48	F:ccCATCGCTGTAAAACTATTTTCC	G/T
7046[3]	F:GCTAGTTCTGATGGTGAAAG R:TGGGAAATTGAAAGGGAAAC	160	53	F:gAACGACAGTATTCAAATTGTGGC	C/T
9699	F:CTTATTGTTGTAGATCTCAC R:TGTGCTTACAATCCATGTCC	142	45	F:TGTATTACATAAAATGGGTAGATA	C/T
3432[1]	F:CACCATTCCACGTTACAAGC R:CGACAGAAGTCAAAACTGAG	117	50	F:AACAAATTCTCACTCTACCATAAAAT	A/G
2271[1]	F:GCGAGAGGTCATTAATTACG R:TGGAACCTCGGGAAATTATC	167	53	F:GGTCATTAATTACGAATTGACGTTGT	A/T
7278	F:TGACAGTTGTCTTCAACAGG R:CTAGTGATTTCTTGGACCTG	156	45	F:acTTAACATTTTTGTTGAAATTTTTAAG	C/T
4655[1]	F:CTTGCAGAACGCATCTGAAC R:GGCAATAAATCGTCAATGGC	314	50	R:TCGGAACGGCCAAAC	T/C
3104[1]	F:GCTGATGCAATGAGTACAAC R:ATCAGTTACTGCTGCTCGTG	323	45	R:TCGCGGAAAACTAGG	T/A
AG-0383F[1]	F:ATAAGAAGCCTGGGCTCCAC R:AGACACAATTTCACCCCCAG	357	48	F:GGCTCCACCAACCTAA	A/G
4655[3]	F:CTTGCAGAACGCATCTGAAC R:GGCAATAAATCGTCAATGGC	314	49	R:aaACTGGGCATGCAGTC	G/A
4655[2]	F:GGCAATAAATCGTCAATGGC R:CTTGCAGAACGCATCTGAAC	314	50	F:CTGAGTGTGACCAGTGC	A/G

Table S4.3 (Continued)

5368[2]	F:AGATGATCTCTGTTTCCGGTC R:CTTCAGAGGTTGTAGTCAA	352	47	R:TGTTCCTACACGTTTCAA	C/A
4913[1]	F:CTTTCCTTCCTTCTTCCAC R:AGACGTGTTTGTGAAGGAGG	282	47	F:CCTTCTTCCACTTTCACTA	C/A
5368[3]	F:AGATGATCTCTGTTTCCGGTC R:CTTCAGAGGTTGTAGTCAA	352	46	R:TGTTTCCGGTCTTTTTTAG	T/C
3182[1]	F:CAATCGACGTTAATGACGGG R:ATCGTACCAGGAACTAGTCG	302	46	R:gaggGGAGCCGTAGGTCTA	G/T
618[1]	F:CTGACCTTTTGCTCCTTTGG R:GCGGCCTTGATGAATTTTGG	313	46	R:cTGCATACTTTTGTCACTTC	G/A
6579[1]	F:GGAGTGCTGGATGTGTTCTG R:CAACGTTAAACCGACAAGAC	273	45	F:AGGTCTGTAGTATCTTAATCA	A/G
6023	F:AGTCCAACTGTCTGTGAATG R:TCATACACGCTTTCGAACAC	329	46	R:TGCGAATTACGATATGAATTTA	G/A
874[1]	F:GGATGCGAAGGGGATTCAG R:CCATCAAAAACGCCACATCC	292	46	F:gagcAGGGGATTCAGCTTTTAA	C/T
3182[4]	F:ATCGTACCAGGAACTAGTCG R:CAATCGACGTTAATGACGGG	302	45	F:ctcccACTCTGGTACTGTATCAC	A/G
8322[1]	F:TGACAATTCTTCCAGTAGGG R:GCAGGGCCACTGATAATATG	317	47	R:CAATTCCCAGTTTTAATAGTTCT	T/A
3843	F:CACCACGGCAAGTTTATTTC R:TTTGCACTTTGGGCTCCTTG	311	47	R:TCACTACTTTACAAATGATAATGG	G/C
3182[3]	F:CAATCGACGTTAATGACGGG R:ATCGTACCAGGAACTAGTCG	302	48	R:ggtgTCTCGTGCGTGAAATAATTT	G/A
8322[2]	F:TGACAATTCTTCCAGTAGGG R:GCAGGGCCACTGATAATATG	317	45	R:cctCATTCACAGAATAATTTTCCTT	G/A
2271[3]	F:ACTCTTGGGAGGAGTTTCAG R:TAGCCTACAAGTTCCGTGAC	301	45	R:gggTTATCATTTTGTGGTTTCAAAT	T/C
4913[2]	F:CTTTCCTTCCTTCTTCCAC R:AGACGTGTTTGTGAAGGAGG	282	47	F:AACTTATCTTTATTCACTACATTCTC	C/T
6579[2]	F:CAACGTTAAACCGACAAGAC R:GGAGTGCTGGATGTGTTCTG	273	47	R:agTTTACATATATTGCACACGATTTA	T/C
3136	F:ACTTGGGACACTATTGCCTC R:CCTCCGTTTAATTGCCCAAG	298	53	F:ctacGAACAGTAGGAAGAATGCTGGG	A/T
8508	F:CCCAGCAACTAAAGATGCTC R:CAGTCTCATTGACAAAATCC	292	47	R:AGACCTTATTGATCTGTATAAAAATAT	T/A
3182[2]	F:ATCGTACCAGGAACTAGTCG R:CAATCGACGTTAATGACGGG	302	55	F:tccTCTAGGCATATACCGAGGCTACTTT	C/G
2271[2]	F:TAGCCTACAAGTTCCGTGAC R:ACTCTTGGGAGGAGTTTCAG	301	45	F:ggtgTTCAATAGCTTGTAGATTTTAATT	A/G
4679	F:TCTTTCTGTACTTCGAACGC R:CGCAGAACTAACCAAACACG	83	48	R:TCGAACGCTCCAAGT	T/A
AG-0148P	F:TACAAGTTCCACCACCAGAG R:AGTCGATCACGACGAGGTAG	105	52	F:TGGTGGGACGCCAAA	A/T
90	F:ACATTCTGCTGCAGTGTCTC R:TCTTTTGCACGATCTGTTGG	97	48	F:ACTTGCTCCTCCTACG	A/T
3732	F:TGTGAATCGATTGCCAGCAG R:TCCTGGATCTACAGTCCTTC	96	45	F:ccAATGTTAGCGCGTCT	C/T
937	F:CCTAGACCTGTAAATGTGGC R:TCATCACGAGGAGGAAGAAC	97	46	F:GTGGCATTAACTGGAGA	C/T
1978	F:TGATTCGTCACCTTGCAACC R:TCCTGATAACAAAGCTTGCC	110	48	F:cCCTTGCAACCAACATCA	A/G
1032[2]	F:AGCCATTAGCTCCTCTAGGG R:ATTCACTTCCTGCCACAACC	109	46	R:CTCCTCTAGGGAAACATC	G/A
730	F:ACATCTTGATTCCGCTGCAC R:TCTAGGTTATGGGCAGTCAG	98	47	R:aGAGCATGAACTCCACTC	T/C
6771	F:TTTAGAAGTGGCGCAGGAAC R:AAAGTCCCTCCAGTCAAGTC	107	49	F:tTAGTGGGTTGGCAAACT	C/T

Table S4.3 (Continued)

2	F:TTTGCACTTTTGTAACTTG R:AGCATTGGAGTCATACAAGC	102	45	R:agggGCAACAACTTCAGCAT	T/A
2454	F:GGTATCCAAGTAGTCATCCC R:AAGCGGGTTCATGACAAAAG	99	48	F:CCGAGTCACATTATTCATACC	C/T
1847	F:TGGAGACAAGGCCTTCAATG R:AAAGAATCCAATCGCCCTTG	116	49	R:cAGATCGTATACGAAATGGGG	G/A
5136	F:CTGTGAGAACAAAGGTGAAG R:TGGTGTATAATGAGAAGTGC	100	45	R:ttacGGTGAAGCAAAGGAAAT	T/C
5727	F:TAGTTGGATCGCGTAACGTC R:CGACAGGGACAAAACATCAG	97	45	R:CTTCGTCTATTAACAGAAAAAC	G/A
714	F:ATCTAGCCTTGGCAAGTACG R:ACTTGGGTTCTGAGTACATC	97	54	R:gTGGCAAGTACGACCTACACAT	G/A
AG-0501F[2]	F:CTCTGACTTCCATTTCTCCC R:AAGCAGCAAATGCAGCAAAG	103	45	R:ccacCTTCAATAGCATCATCCAA	G/A
3433	F:AGCAGAAAATTCATGGGAAG R:GGCAATATGAATTTTGGGAAG	111	46	F:ggaaACATGTCAATGAGCAATTT	C/T
432	F:GGAAGCGTCTTCAGAAGATG R:TTATCCTCATCCTCTTCCGC	95	47	F:GATGAAAAAGATAAGGCAGATAA	A/G
1555[2]	F:TCTACCAGAGCTCCATCTTC R:GTTCACGTTCTCTGTCTTGG	98	47	R:cacccGAGCTCCATCTTCAGATAG	T/A
541	F:CCACTAGTTTCACGTTCTCC R:CTGAAGGTGGTTTTTTGGAC	105	47	R:tTTGATAAATTACTTTACCATCTGG	G/A
4795	F:TTTGGCCCCACAACTACTTC R:AAACGAAATGGTTGATCGGC	101	51	R:gggatGCAGTGCAGCATAGAAATCT	G/A
518	F:CATCTGGTTTGTAGGCTGTC R:CCTCGAGAGCTCATTGTTAG	119	48	F:aaggaAGATCGTGCTGTAAGAAATC	A/G
1724	F:TTCCCCCTTACTTTAGCAAC R:TCCCTATGATGTTCGGTAAC	120	51	F:aagaCACTTCATCTTTACTGTGGCTT	A/G
AG-0308F	F:AGTGAGTGTCGTTAGAAGCG R:CTGCAGATCTTGACTGTCAT	119	49	F:TCTTTAATGAATACAGAAGACTAATCA	A/T
6718	F:CATTCGTGTGGGCTAGTTTC R:ACTTCGATCTACGAACTGCC	98	49	F:cccaTACACTATCAGCTTTTCCGATATA	A/G
1414	F:CCTTGCATCTTTTTCTTGGC R:AGCCTCGTATACCTTCGTTG	92	48	R:ggctgTTTCTTGGCTTTCCTTTTATTTT	G/A
650	F:CCATATATTGGCACAATAGG R:ACGAGAAGCATTCCGTAGTG	110	47	R:agATAGGTAATCTAGTAATACTGTACAA	G/A
3838	F:TCACTTGCGACTTTTGAGAC R:GACTATTTTGTGAGACGTAG	113	45	R:agaAGACATTGTAAAATTTAAAATGATC	T/C
AG-0159F	F:ATGTTTTGAGACAGACCCCG R:TCTGGTAAGGGAACTTGTCG	118	49	F:GACCCCGACACCAAA	C/T
1177	F:AGTCACCCACAGATTAAGCG R:ACACTGAGTTCCTTCTTTGC	100	52	F:AGCAAGGCATCGCCA	G/T
1903	F:AACCACTCGCTAGTGATGAC R:GGACGATATGGAGGATTTGG	117	55	F:GTGCTGCGAGGGGGT	G/T
3268	F:TCTATCTTCCACCCACACAC R:GTTTCTCAGCTGGAAAATAC	116	46	R:CCCACACACCAGTATG	T/C
AG-0334P[1]	F:TTGATTTTCTCGCGAGAGCC R:CCTCTCTTACTTGTTACAGG	106	50	F:aCGAGAGCCGGAGTCT	A/G
855	F:GCATTGGGAGTAAGCAATCG R:TAAGCGACTGCTAGATACCC	91	46	F:CGTATTCCCGATGTTCT	A/T
5510	F:TGTCAACTGTGTGAAGCGTG R:GACAAACTTCTACGAGACCC	100	48	R:gGCGTGTTTTGTGAGAGA	C/A
4914	F:GACACCTGAACTGTTCAGAC R:CCAGTGTTCTTGAGTTCAGC	97	46	R:ACTGTTCAGACAATTTTCC	G/A
280	F:GGTGCTGTTTGAGAGTTGAC R:TCAACGACACAGGTTGTTGC	109	51	R:GAGCAGCCAGTGGTTTTAA	G/A
1341	F:GGTTTGCCAACAAATTGTCC R:GCTTTTGCAGATTCAATGTAG	108	47	F:tCAACAAATTGTCCACACAT	C/T

Table S4.3 (Continued)

1275	F:TGGAACTCGAGTAGCTCTTG R:TACACAAGAGGGTCCACTTC	98	47	F:gGAGTAGCTCTTGACATGAC	A/G
1415	F:GAGATTTTTTGTGAGGTAG R:GTCACCTGAATTTTCAGAAC	112	49	F:tgTTTGTGAGGTAGTGCAGT	A/G
1305	F:GGAGAATTTCTTCTTCTTT R:CCCATATCATCTCAAGTTCT	120	45	R:ttcgGCTTTGCAAAGCTCTTA	T/A
861	F:ACGGTTTTGGATGTGTGTCG R:AGCTGTCACATCAGTTCCAG	104	49	F:TAGAGAAAGTGGAATGAAGGT	C/T
989	F:AGCCCAAACTACCTTGGAAC R:CTGGAGGACAATTGCTGGG	115	45	F:cggATACATTCCCACCATCTAT	A/G
5777	F:AGAGGTTGGTGGAAAACGTG R:GTAGGCACAATGAAAGTCTG	110	51	F:AAACGTGGATATATGCCCAAAT	C/T
2833	F:CATTGGGATCTGGCATTGTG R:AGTATGACACAGACTTCTAC	98	50	F:ccGGCATTGTGTAGAAAGGTCT	C/T
4450	F:TCATTTGTTCCACGTACATC R:TTTCTGTTCTTTTCTTGAC	117	47	R:tgggACATTTAATCATCCCCACA	G/A
8229	F:GAGGACATTTTACGAGCAGG R:ACTCAACGGCTAATTGGCTG	97	45	R:ggtcTTGAATCGATGGGATAGTA	G/T
6271	F:AATAATCAGCAGTTTGTCG R:GCGCAAAACTAATTGCTGTC	118	52	F:taagCAGCAGTTTGTCGTCTTCTG	C/T
580[1]	F:CGGTTAATTACCGTCTGTCG R:CAATCTCAAAGTGGCTGTCG	97	47	F:aagagCGTTTACGCTTGACATTTT	A/G
AG-0115F	F:CCGACGTAAAAGTACTTAGTG R:GTTTACCCAACTTTCGATCAG	116	46	R:cccgTCTTTTAGCAACACTTTCTTA	G/A
3758	F:GAAGAACATAATCCATGACC R:ATGGGAATAGTTTTAGGTGG	112	47	R:ccTGACCAATTATCATCAAATAACA	G/A
7083	F:AAGAGGAGCGGAAGAGAATG R:TTTGGGTGAAAGAGAGTGGC	120	57	F:aggtTTGTTGGATGCAAGTCACGCC	C/T
6557	F:ATGTCATCATTCTCTTGAGC R:GACTACTATAATAATTGTAC	107	45	R:tacaaTCCAATTGAATAGAATTCACT	G/A
3566	F:GATCTTCTGCAAGCTTTGGG R:TTCAGTCCATGTTAGCAAAG	102	50	F:ggtaGTAAAGATCTTCAGGACAGAGT	C/T
1372	F:CGCAACAATCATCACAATCG R:GAGTACGTGTTCTCGATGTG	93	47	R:gtcATCATCACAATCGTATTATACAAT	G/T
1145	F:GTCAAGTTCCACAGAATCAC R:TTCTACCTGCTGCTGCAAAG	118	50	R:ATCACTTAATATTTTCATGTGAGATGT	T/C
6026	F:CTTTAAGTGCTGGCAGCAAC R:CGCAAGTCATAGCATATTCC	99	52	F:tgACATTTTTTGACACACCATTTTGTAC	C/T
17333[2]	F:TGTTTTAAGACATTCCTGCC R:TGAAAAAGTAATATTCCTTG	118	45	R:acctgTGTATCTAATAGCAATCCTTTAT	T/C
3432	F:TCCAGGCATCAGAATTCCTC R:GCAATGCAGAAATCCTTCGG	119	56	F:ggggaAGGCATCAGAATTCCTCTGCAAC	C/T
1667	F:ACCTCTGCCCCTCGAGACA R:CTGAAAGATGAGAGGCCAAG	100	52	F:ACCCTGCCTCTTGCC	A/T
7164	F:CTTCACCAACAACAGACCTC R:ATATCCAGGCAGAAATTCCC	110	47	R:TGGTGAACGCCAATG	T/A
625	F:GGTTTCCTCTCCTTCTTCAG R:GGAGGAACAAGCAAGTGAAC	112	46	R:gACCACGCAACCTTTT	G/A
AG-0099F	F:ACGTGATCTGGCAAATGGAG R:CCCAAGCCAAAATATGCGAG	102	55	F:TCTTGGCGGATGCCCC	C/T
1121	F:TCTCCTCCAGCTAGGTTTTC R:GTCACATAAGAGTTCAGAAAG	110	47	R:AACCACAGGTGTTGTAG	G/A
14937	F:GAGGTCTTGCAATGATGCAC R:AAGACGCGTGATTAAGAGAC	105	49	R:ggGCTGGATCGATGCCT	T/C
6163	F:GGTTCCAGCAGAGAAAGTAG R:CCACACTCGTACACTTCAAC	104	46	R:CAAAGCACTTCATCTCAC	T/C
AG-0023F	F:CTCTCCTTTTCCTCTTGGAC R:CTTCCAGAGGCAACTCAATG	107	47	R:CCTCTTGGACAACTGTTT	C/A

Table S4.3 (Continued)

3528	F:GAAAAAGAAGGCGGAACTGG R:TGTCCTTGAGACTGAAGTGC	94	47	F:GTTGCTGATGGATGAAGA	A/G
2292	F:TCAGCAAGTGATCTCCTGTC R:GCAGTATCTGGACCTAGTTC	106	46	R:cTGTTAAGTGCGTACAGTA	G/A
5120	F:GTTGAACTTTCTCGTTCTTC R:CTTATTGCGTCGTCTGCTTG	88	46	F:TCTTCATCTAGTGTTTCTCT	G/T
2040	F:TGTTGTGCAGCATTTGCATC R:GTCTGTTTTTTGCTGTCGGG	110	48	R:cccGCATTTGCATCATCTCCT	C/A
755	F:TCTCAGCGAATCCATCAGTC R:AGGATGAGTTTAGCAGCTCC	100	50	F:gTCCATCAGTCTGAAGCTTAC	C/A
AG-0005F[4]	F:AGAGGCCCAAAACCTGGAAC R:GAAATAGCCCGTTTGTTGCC	100	46	F:ACGAGTATGTTTTATCAAGAC	G/T
1774	F:CAACCTGACCAAGAATCCAC R:CTTGTCCCTTTTGATCCCAG	94	49	R:ggccTGGATCTCCTTTGGTTCC	G/A
2182[2]	F:ACACATCTTCAATGTTCTGC R:AGTTTACACATGCTCTGCTG	99	47	R:atgtAACATGATTGTGGTGACA	G/A
543	F:ACTGGAACACCATCCTCAAC R:GCTCCCCAAAATGTTCATCC	100	46	R:TTTTCCATAGTAGTCCTAATTTT	G/A
1638	F:CAAATTTGTGTGTGTGGCCC R:GAGTCTTTGGCACTGATCC	90	51	R:cctaCCGAAATCAGGATGTCCAC	T/C
5067	F:ACTTGGAATCTTCAGCTGAC R:CAGTTGAATCGGATTGTTCAG	98	45	F:gggaAGCTGACTTCTTATGTTTG	A/G
5711	F:GCCAAGTGAGAGTAAAGTTC R:TGATGAACCCTTTGGACTGG	100	51	R:tcGTTCTGTTCACCTTTTCTCCTT	C/A
5021	F:ACTTAACGCGAACTGTTTTC R:GTGCTTGGGCCTTTAATCAG	100	47	R:taaaGCGAACTGTTTTCTCTTATG	T/A
3084	F:CACGCCATGGCATCATATTC R:TGAGGTACTGAACGAAGGTC	94	46	F:cccccGCATCATATTCTGAAGATCG	C/G
367	F:ATGGCACTGAGTCTGTGATG R:ATACAAGACAAACACAAGC	118	46	F:AAATTCATGAAACAATAAATTTGTG	A/G
1101[1]	F:GCGCGTTGAGTTTGAAGAAG R:ATCGGCGGACAATGAATTGC	100	48	F:gtgcAGTTTGTGTTACAGTAGTTGA	A/G
5052	F:CGCTCCATTATTGCATCTTG R:TCTGCTAGAGTTCGAGCTTC	101	50	F:tatttCATCTTGCTGGACTTTCTGTA	A/G
2100	F:GAGCTCGTTATGTAGGTTCC R:GAGCATCTTTAGTTGCTGGG	110	55	R:gggtTAGGTTCCATGGTTGCAGATGT	T/C
1369[2]	F:GGGCCATACTCTTTGCAAAT R:TTCCTATATAGTAATTGGC	107	50	R:TTGCAAATATCATCAGTGATATTATCC	G/A
1313	F:GGGTTCTTCATTACTAAAGG R:ACCACTGTTCCAATTTGAAG	118	47	F:ccgtaTTCATTACTAAAGGTGCATTTT	A/G
2467	F:AATTCAGAAAGTCCCATGCG R:TATTGCCTGCATCTTATGTG	106	59	R:cccacAGTCCCATGCGCTTTTTGTGCTC	T/A
8026	F:TCTTCTCTCCTCCTCCTCAC R:CCCTCTCTTCAAGCCTTTTG	105	50	F:ACGCGCAAGACGTTC	C/T
AG-0313F	F:ACTTTTGCACCTTCTCCACC R:ATGCAGTTGATTACGCGGTG	98	48	R:GCACCTCCTCTACCAC	T/A
3968	F:TACTTTTGAATCGGCCTCGC R:ATATGCTCGCTCTGAGCTTC	109	47	R:gGGCCTCGCACATAATC	T/C
1392	F:AAGAGTTGCAACGGCAAAGG R:TGCTTCCAAATCCACTGCTG	101	46	F:AAGAAGGGGATGAAAGG	C/T
AG-0005F[5]	F:TGTTTAAAGAGCCCGACCAC R:AAGGAATGGTAGGACGGTTG	111	60	F:aACCCCTCCTGGCCAGCA	A/G
1309	F:TCCTGTCTTCAGCATCCAGC R:GTGAATGAGGACTTAGTTGG	103	45	R:ATGAAATTCTTGCAGTCC	T/C
80	F:GTGACTGCAAGGTCATTCTG R:CACTGGCAGGAGCATTCTTC	97	50	F:GCAACTGTTGAAGCTTGG	C/T
9851	F:GGTTTCTTTCTATGATCATC R:AAACGTGCTAATAGCAGCAG	112	50	F:TGGAACTGGAACTATCAGC	A/G

Table S4.3 (Continued)

1721	F:AAACAGAACTTCTGACACCG R:CCCTGAATTTGTCACGAAGC	100	46	F:cTTGTGAAAAATCTTCCTCC	A/T
2887	F:GCGATTGTATCGATCTGAGG R:GAAGTGTCGAGGCATCAAAC	93	47	R:aagCGATCTGAGGCACAAAT	G/T
136	F:AGAAGGCGGCGAAGATTAAC R:ATGTGAACAATCCTCGCCTG	107	47	F:tGCGAAGATTAACAGAGTGA	C/A
AG-0076F	F:TGCTAAAGCATCATCTCCG R:AGCACGTGCCATCGCGAAT	99	51	R:tAGCATTGGAGTTGGCTAAAG	G/T
4411	F:GTTTCCAACTAAACGTGTTCC R:CTGCTATTCGAAAACACGTTC	119	46	R:tacCTAAACGTGTTCCAAAGAG	G/A
AG-0501F[3]	F:AAGCAGCAAATGCAGCAAAG R:CTCTGACTTCCATTTCTCCC	104	45	F:tcccTGCAGCAAAGTTAAAAGA	A/G
87	F:GTCAGTGGGATTTAGATGATG R:TCTTTTTTAACAGGTTCAGC	114	47	F:ATGAAAAAGATGAAGAACTGAC	A/T
1790	F:AAGAGAGCCATACTCTTCCC R:ACCCATGAAACTTGGATGAG	113	45	F:gAATGAGGAAAACAAAGAAATAAT	C/T
267	F:TCCAAGAGTTGATGACCCTG R:GTCCTTTTTATGTTGATCTGC	90	49	R:ggagGGGAAACTCTGGGAAAAATT	G/A
1699	F:CACACCGCATATGTAAGTTC R:GGCAAAGTAAACATGCATTCC	109	45	R:ggtATATGTAAGTTCTATTTCCTGG	T/A
99[1]	F:GTGATCTCTTTTCGTGTGGC R:GATCCAAGAAAATCCCTTGAG	119	45	R:aaACATATTATATAACTTTCAACAGC	G/T
1926	F:TGGGCAAAATAATATACTGAC R:CAGACACGAATGAGCAATTA	119	46	F:CAAAATAATATACTGACATAACGATAT	G/T
2570	F:ATAGCCAATATAGTGTCTG R:AACATGAAGTTCTATACAC	102	45	R:ctccTCCAGAAATGTTTTATTTATCTAC	T/C
4328	F:GGTGGTTCATTATTATCTTG R:CAATCAACGCCAAAATCTTC	105	45	F:ATCTTGATTAATATTTAATACTCCTAAT	A/T
6128	F:CAAGATACATTTCTTGTGAGC R:AATCAATAACAGTGATGTC	99	46	R:AATAAATAAATAAACAAAACTTGTATGA	T/A
AG-0005F[1]	F:CTCGCAACAAGATGGACTAC R:TGGGATTGAACCCCCATAAG	97	49	F:CTTTTCCCCGTGGTACA	C/T
AG-0005F[2]	F:CACCTTGTTGGCCAGGATTG R:ACCACTCCGAAACCCCGAC	119	48	R:CCAGGATTGGCTAAGGA	G/C
AG-0005F[3]	F:ATCACAGCCCTACAACCAAG R:AATGCTCGTCCTCGTAATCG	115	52	F:CTACAACCAAGTAGCCCCT	C/T
AG-0203P	F:TTTGCAAAAGCCTCCTTCCG R:TTGGTGCTACGTTGCTACTC	118	46	F:AACCAAACTGATGAAGAAG	A/G
AG-0334P[3]	F:CTGAAACCTCCCAAATACCC R:TTGGCAACCAGAATTGCAGG	110	48	R:TTTTTTCGCACCTCTTTTTT	G/A
1147[1]	F:CTTTTAAACGGTGCACTCGC R:GAGGTGGCTATGTTTTGTGG	92	46	F:caAAAACCAAGCATATGTCAA	C/A
99[2]	F:GTGATCTCTTTTCGTGTGGC R:GATCCAAGAAAATCCCTTGAG	119	47	R:GCATATGTAATTGAGAACACA	G/T
1101[2]	F:CTATTACCTGGGTGGTTTGG R:TTTTTTCAGGCGGCATGGAC	109	47	F:AGGAATACTAGGAGAATTGTC	A/G
AG-0254P	F:GCCGTATTCATTACATCGGG R:TGGACTTGGACGAGAACAAC	99	45	F:TCCTCATTCATCCTAATAAAAT	G/T
14713	F:CCGGATAGCATGTTCATCTC R:GAATGCGTCGAAATGGGATG	99	47	R:ggATCATGCAGTTCATTATCTCT	T/C
580[2]	F:GTATATACTGATCCTGATTG R:GATGCATGGACATTTTAACTG	110	50	F:ACTGATCCTGATTGAATGTGATA	C/T
AG-0501F[4]	F:CGAACAAATCCACTGTAATCC R:AGCTGCCTCAACCTTTTTCC	98	50	R:TCCACTGTAATCCATCAATTCATA	G/A
AG-0334P[2]	F:TTGATTTTCTCGCGAGAGCC R:CCTCTCTTACTTGTTACAGG	106	61	F:ATTTTCTCGCGAGAGCCGGAGTCT	A/T
8354	F:CAATCTACAAGTCATTAGAGG R:ATGCACACTTTCCATCCAGC	114	45	R:cacAAAATAGTTTTCTCTGTTTCTG	T/A

Table S4.3 (Continued)

6057	F:CACGATAAAATGGTTACAG R:CGTGGTGGAAGTTTGTTCTG	118	48	R:AAATGGTTACAGTTTAACAAAAATG	G/T
6128	F:AATCAATAACAGTGATGTC R:CAAGATACATTTCTTGTGAGC	99	51	F:ggaAATAACAGTGATGTCTACTGCAA	A/T

bp: expected amplicon length in base pairs; T: amplicon melting temperature
Contig numbers represent different SNPs assayed per contig
All forward and reverse amplicon primers included an IPLEX chemistry primer tag (ACGTTGGATG)
SNP indicates the targeted nucleotide base that is incorporated at the end of the extend primer
Bold lines indicate the division between the nine assays

Table S4.4 Allele frequencies in allopatric populations of *G. pennsylvanicus* and *G. firmus* for 154 SNP markers successfully genotyped.

	G. pennsylvanicus				*G. firmus*				
Locus	**ITH**	**SCO**	**SCR**		**GUI**	**PAR**	**TRI**		***D***
2	1	1	1	**1**	0	0.083	0.167	**0.083**	0.917
80	1	1	1	**1**	0.083	0.042	0.333	**0.153**	0.847
87	1	1	1	**1**	0.042	0.042	0.292	**0.125**	0.875
90	1	1	1	**1**	0.167	0	0.125	**0.097**	0.903
94	1	1	1	**1**	0.167	0.083	0	**0.083**	0.917
99	1	1	0.909	**0.971**	1	1	1	**1**	0.029
136	1	1	1	**1**	0.125	0.042	0	**0.056**	0.944
211	0.958	1	1	**0.986**	0	0.042	0	**0.014**	0.972
280	1	1	1	**1**	0.417	0.375	0.250	**0.347**	0.653
367	1	1	1	**1**	0.083	0	0.042	**0.042**	0.958
402	1	1	1	**1**	0	0	0	**0**	1
425	1	1	1	**1**	0	0.042	0	**0.014**	0.986
432	1	1	1	**1**	0.042	0.167	0.042	**0.083**	0.917
518	1	1	0.909	**0.971**	0	0	0	**0**	0.971
541	0.917	1	1	**0.971**	0.042	0.042	0.125	**0.069**	0.902
580	1	0.958	0.909	**0.957**	0	0	0	**0**	0.957
618	1	1	1	**1**	0	0	0	**0**	1
625	1	1	1	**1**	0.333	0.333	0.208	**0.292**	0.708
650	1	1	1	**1**	0.333	0.667	0.417	**0.472**	0.528
714	0.958	0.958	1	**0.971**	0.292	0.250	0.375	**0.306**	0.666
726	1	1	1	**1**	0	0	0	**0**	1
730	1	1	1	**1**	0	0.125	0	**0.042**	0.958
755	1	1	1	**1**	0.083	0.167	0.083	**0.111**	0.889
827	1	1	1	**1**	0	0	0	**0**	1
855	1	1	1	**1**	0.250	0.042	0.167	**0.153**	0.847
874	1	1	1	**1**	0	0	0	**0**	1
937	1	1	1	**1**	0	0	0	**0**	1
963	1	1	1	**1**	0	0	0	**0**	1
989	1	1	1	**1**	1	1	1	**1**	0
1032	1	1	1	**1**	0.250	0.083	0.042	**0.125**	0.875
1101	1	1	1	**1**	0	0.083	0	**0.028**	0.972
1121	1	1	1	**1**	0	0	0.208	**0.069**	0.931
1145	1	0.792	1	**0.929**	0	0	0	**0**	0.929
1147	1	1	1	**1**	0	0	0	**0**	1
1177	1	1	1	**1**	0	0	0	**0**	1
1231	1	1	1	**1**	0	0	0	**0**	1
1234	1	1	1	**1**	0.042	0.042	0.250	**0.111**	0.889
1275	1	1	1	**1**	0.250	0	0.042	**0.097**	0.903

Table S4.4 (Continued)

1305	1	1	0.818	**0.943**	0	0	0.250	**0.074**	0.869
1309	1	1	1	**1**	0.042	0	0.125	**0.056**	0.944
1313	1	1	1	**1**	0	0	0.042	**0.014**	0.986
1341	1	1	1	**1**	0	0	0.042	**0.014**	0.986
1369	1	1	1	**1**	0	0	0	**0**	1
1372	1	1	1	**1**	0	0.083	0	**0.028**	0.972
1374	1	1	1	**1**	0	0	0.083	**0.028**	0.972
1412	1	1	1	**1**	0	0	0	**0**	1
1414	1	1	1	**1**	0.042	0	0.250	**0.097**	0.903
1513	1	1	1	**1**	0	0.083	0	**0.028**	0.972
1539	1	1	1	**1**	0	0	0	**0**	1
1555	1	1	1	**1**	0	0	0	**0**	1
1638	1	1	1	**1**	0.292	0.417	0.250	**0.319**	0.681
1667	1	1	1	**1**	0	0	0	**0**	1
1724	0.875	0.958	0.909	**0.914**	0	0	0	**0**	0.914
1774	1	1	1	**1**	0	0.042	0	**0.014**	0.986
1790	0.917	0.958	0.909	**0.929**	0.083	0.083	0.083	**0.083**	0.845
1847	0.875	0.917	0.864	**0.871**	0.083	0.125	0.167	**0.111**	0.76
1851	1	1	1	**1**	0.042	0.083	0.042	**0.056**	0.944
1978	1	1	1	**1**	0.250	0.250	0.250	**0.250**	0.75
2100	1	1	1	**1**	0	0	0	**0**	1
2182	1	1	1	**1**	0	0	0	**0**	1
2271	1	1	1	**1**	0	0.083	0.167	**0.083**	0.917
2292	1	1	1	**1**	0.500	0.500	0.833	**0.611**	0.389
2361	1	1	0.955	**0.986**	0.083	0	0.042	**0.042**	0.944
2454	0.958	1	0.909	**0.957**	0.250	0.667	0.625	**0.514**	0.443
2467	1	1	1	**1**	0	0	0	**0**	1
2570	1	1	1	**1**	0	0.083	0	**0.028**	0.972
2733	1	1	1	**1**	0	0	0	**0**	1
2831	1	1	1	**1**	0	0.083	0.167	**0.083**	0.917
2833	1	1	0.909	**0.971**	0	0	0	**0**	0.971
2864	1	1	1	**1**	0	0.083	0	**0.028**	0.972
2989	1	1	1	**1**	0	0	0	**0**	1
3084	1	1	1	**1**	0.042	0	0.167	**0.069**	0.931
3136	1	1	1	**1**	0.083	0.083	0.292	**0.153**	0.847
3182	1	1	1	**1**	0	0	0	**0**	1
3268	1	1	1	**1**	0	0.042	0.042	**0.028**	0.972
3344	1	1	1	**1**	0	0	0	**0**	1
3422	1	1	1	**1**	0	0	0.083	**0.028**	0.972
3432	1	1	1	**1**	0	0	0	**0**	1
3433	1	1	1	**1**	0.125	0.292	0.250	**0.222**	0.778

Table S4.4 (Continued)

3528	1	1	1	**1**	0.083	0	0	**0.028**	0.972
3555	1	1	1	**1**	0	0	0	**0**	1
3566	1	1	1	**1**	0.292	0.042	0.375	**0.236**	0.764
3732	1	1	1	**1**	0.292	0.042	0.333	**0.222**	0.778
3838	1	1	1	**1**	0	0	0	**0**	1
3843	1	1	1	**1**	0.042	0	0.042	**0.028**	0.972
3968	1	1	1	**1**	0	0	0	**0**	1
4205	1	1	1	**1**	0	0	0	**0**	1
4328	1	1	1	**1**	0	0.500	0.250	**0.417**	0.583
4361	1	1	0.955	**0.986**	0	0	0	**0**	0.986
4450	1	1	1	**1**	0	0	0	**0**	1
4481	1	1	1	**1**	0	0.500	0.250	**0.417**	0.583
4655	1	1	1	**1**	0	0.042	0.042	**0.028**	0.972
4679	1	1	1	**1**	0.042	0	0.125	**0.056**	0.944
4913	1	1	1	**1**	0	0	0	**0**	1
5021	1	1	1	**1**	0	0	0	**0**	1
5052	1	1	1	**1**	0	0	0	**0**	1
5120	0.875	0.833	0.909	**0.871**	0.292	0.042	0.250	**0.194**	0.677
5131	1	1	1	**1**	0	0	0	**0**	1
5136	1	1	1	**1**	0.083	0	0	**0.028**	0.972
5177	1	1	1	**1**	0	0	0.208	**0.069**	0.931
5214	1	1	1	**1**	0	0	0	**0**	1
5368	1	1	1	**1**	0	0.042	0.042	**0.028**	0.972
5510	1	1	1	**1**	0.125	0.458	0.333	**0.306**	0.694
5556	1	1	1	**1**	0.042	0.167	0.042	**0.083**	0.917
5711	0.917	0.958	0.909	**0.929**	0.083	0.083	0.083	**0.083**	0.845
5727	0.917	1	0.955	**0.957**	0.083	0	0.083	**0.056**	0.902
5777	1	1	1	**1**	0	0	0	**0**	1
5961	1	1	1	**1**	0	0	0	**0**	1
6023	1	1	1	**1**	0.042	0.042	0.125	**0.069**	0.931
6026	0.917	1	1	**0.971**	0.042	0	0.125	**0.056**	0.916
6030	1	1	1	**1**	0	0.042	0	**0.014**	0.986
6057	0.958	1	0.909	**0.957**	0	0.083	0	**0.028**	0.929
6128	0.958	1	0.909	**0.957**	0	0.042	0.125	**0.056**	0.902
6271	1	1	1	**1**	0	0	0.208	**0.069**	0.931
6557	1	1	1	**1**	0.042	0	0	**0.014**	0.986
6571	1	1	0.909	**0.971**	0	0.042	0	**0.014**	0.958
6579	1	1	1	**1**	0.042	0.083	0.083	**0.069**	0.931
6718	1	1	1	**1**	0	0	0	**0**	1
6771	1	1	1	**1**	0	0	0.208	**0.069**	0.931
7046	0.917	0.875	0.955	**0.914**	0	0.083	0.042	**0.042**	0.873

Table S4.4 (Continued)

7083	1	1	1	**1**	0.292	0.083	0.292	**0.222**	0.778
7153	1	1	1	**1**	0	0.500	0.083	**0.194**	0.806
7164	1	1	1	**1**	0.083	0	0	**0.028**	0.972
7469	1	1	1	**1**	0.125	0.083	0.125	**0.111**	0.889
7566	1	1	1	**1**	0.125	0	0.250	**0.125**	0.875
8026	1	1	1	**1**	0	0	0	**0**	1
8229	1	1	1	**1**	0	0	0	**0**	1
8257	1	1	1	**1**	0	0	0	**0**	1
8322	0.958	1	0.955	**0.971**	0	0	0	**0**	0.971
8354	0.917	1	1	**0.971**	0	0	0	**0**	0.971
8354	1	1	1	**1**	0	0	0	**0**	1
8375	1	1	1	**1**	0	0	0	**0**	1
8612	1	1	0.955	**0.986**	0	0	0	**0**	0.986
9839	1	1	1	**1**	0	0	0.250	**0.083**	0.917
9851	1	1	1	**1**	0	0.042	0.042	**0.028**	0.972
10368	1	1	1	**1**	0	0	0	**0**	1
11695	1	1	1	**1**	0.042	0	0	**0.014**	0.986
12397	1	1	1	**1**	0	0	0	**0**	1
14713	1	1	1	**1**	0	0	0	**0**	1
14741	1	1	1	**1**	0	0.333	0.750	**0.361**	0.639
14937	1	1	1	**1**	0	0	0	**0**	1
16015	0.958	0.917	0.955	**0.929**	0	0.125	0.167	**0.083**	0.845
AG-0005F	0.917	0.958	0.455	**0.814**	0.208	0.042	0.833	**0.139**	0.675
AG-0023F	1	1	0.955	**0.986**	0.125	0.125	0.042	**0.097**	0.888
AG-0099F	0.500	0.455	0.455	**0.500**	0.250	0.292	0.583	**0.681**	0.181
AG-0115F	0.750	0.458	0.455	**0.557**	0.500	0.750	0.417	**0.557**	0.002
AG-0148P	1	1	1	**1**	0.292	0.042	0.042	**0.125**	0.875
AG-0159F	1	1	1	**1**	1	1	1	**1**	0
AG-0254P	1	1	1	**1**	1	1	1	**1**	0
AG-0308F	1	1	1	**1**	1	1	1	**1**	0
AG-0313F	0.583	0.750	0.682	**0.671**	0.542	0.250	0.417	**0.403**	0.269
AG-0334P	1	1	1	**1**	0.250	0.750	0.083	**0.361**	0.639
AG-0383F	1	1	1	**1**	0.042	0	0.042	**0.028**	0.972
AG-0501F	1	1	1	**1**	0	0	0	**0**	1

Table S4.5 Sequence descriptions and functional annotation from Andres et al. (2013) for 125 SNPs used in genomic cline analysis. Included are the number of SNPs per contig (SNPs), the number of amino acid replacement SNPs per nonsynonymous site relative to the number of silent SNPs per synonymous site (pN/pS), sequence description, functional annotation using BLASTX and genomic clines analysis deviation category. Contigs in bold indicate markers that had significantly reduced introgression within the hybrid zone.

Contig	SNPs	pN/pS	Sequence description	Function	Dev.
2	3	-	Translation factor *SUI1*-like protein [*Aedes aegypti*]	Translation initiation factor	P → F
80	14	0.175	Eukaryotic translation initiation factor 3 (*eIF-3)*-like protein [*Nasonia vitripennis*)	Translation initiation factor	het +
87	3	0	Nucleosome assembly protein 1 (*NAP-1*)-like isoform [*Nasonia vitripennis*]	Histone binding	ns
90	2	0	Ribosomal protein L13 [*Xenopsylla cheopis*]	Ribosome structural constituent	P → F
94	**7**	**∞**	**Similar to *Gryllus bimaculatus* mRNA GB18903**	**NA**	**het -**
136	7	∞	Mn-Superoxide dismutase (*Sod2*) [*Bombyx mori*]	Metal ion binding	P → F
211	**2**	**∞**	**Translation factor *SUI1*-like protein [*Aedes aegypti*]**	**Translation initiation factor**	**het -**
367	3	0	Similar to *Tribolium castaneum* putative signal peptidase 12kDA subunit	Peptidase activity	ns
402	**1**	**-**	**RNA binding motif protein 4 (*Rbm4*)-like protein [*Apis mellifera*]**	**RNA binding**	**het -**
425	2	0	COP9 signalosome complex subunit 4 (*SGN4*)-like protein [*Nasonia vitripennis*]	Protein binding	ns
432	11	0.842	Similar to *Tribolium castaneum* LOC398543 protein	Nucleic acid binding	ns
518	2	-	Leucine zipper and ICAT homologous protein [*Suberites domuncula*]	Beta-catenin binding	F → P
541	3	0	Similar to *Tribolium castaneum* LOC660605 protein	NA	het +
580	3	-	Nuclear transporter factor 2 (*NTF2*)-related export protein [*Nasonia vitripennis*]	Protein transmembrane transporter activity	ns
618	5	∞	Conserved protein (similar to Cyclin-D1-binding protein 1)	Binding	ns

Table S4.5 (Continued)

726	4	∞	Similar to *Gryllus bimaculatus* mRNA, GB01128	NA	ns
730	4	-	Conserved hypothetical protein [*Culex quinquefasciatus*]	NA	P → F
755	6	0.612	Apoliopophorins precursor [*Pediculus humanus corporis*]	Lipid binding	P → F
827	4	0	Aspartate aminotransferase [*Anopheles gambiae*]	Pyridoxal phosphate binding	ns
855	4	0.889	Similar to *Tribolium castaneum* AGAP003463-PA	NA	P → F
874	**5**	**-**	**ES1 protein homolog, mitochondrial-like [*Macaca mulatta*]**	**Methyltransferase activity**	**het -**
937	3	∞	Dynactin subunit 4 (*Dctn4*) [*Aedes aegypti*]	Protein N-terminus binding	P → F
963	**3**	**-**	**Similar to growth hormone-inducible soluble protein [*Ixodes scapularis*]**	**NA**	**het -**
1032	1	0	Ubiquitin-like modifier activating enzyme 5 [*Acheta domesticus*]	Co-factor binding; binding	ns
1101	**3**	**0.130**	**Endopeptidase inhibitor-like protein [*Phlebotomus duboscqui*]**	**Peptidase inhibitor**	**het -**
1121	**2**	**0.306**	***Rab11* interacting protein**	**GTPase activity**	**het -**
1145	2	0.292	Ribophorin putative protein [*Ixodes scapularis*]	Transferase activity	F → P
1147	**3**	**0.630**	**ERGIC-53-like protein (*LMAN1*) [*Tribolium castaneum*]**	**Protein binding; metal ion binding**	**het -**
1177	4	0.288	DDB1- and CUL4- associated factor 13 (*Dcaf13*)-like protein [*Bombus terrestris*]	Molecular function (catalysis or binding)	P → F
1231	4	0.917	Glutathione S-transferase [*Gryllotalpa orientalis*]	Transferase activity	ns
1234	1	-	Notch-like gene [*Drosophila simulans*]	Chromatin binding; protein binding; receptor activity	ns
1275	3	∞	Protease regulatory subunit S10B [*Culex quinquefasciatus*]	Peptidase activity; ATP binding	P → F
1305	**2**	**0.293**	**Twinfilin-like protein [*Apis mellifera*]**	**Actin-binding; protein tyrosine kinase activity**	**het -**
1309	4	0	Similar to *Tribolium castaneum* B52 CG10851	NA	het +
1313	9	0	Pacifastin light chain-like protein [*Culex quinquefasciatus*]	Peptidase inhibitor activity	P → F

Table S4.5 (Continued)

1341	5	0.076	Citrate lyase beta-like protein [*Tribolium castaneum*]	Lyase activity	ns
1369	3	-	*Acyrthosiphon pisum* outer dense fiber-like protein	Sperm tail structure	F → P
1372	**3**	**0**	**Lambda-crystallin homolog (*CRYL1*) [*Nasonia vitripennis*]**	**NAD+ binding; protein homodimerization activity**	**het -**
1374	3	∞	Conserved protein: unknown	NA	ns
1412	4	-	Insect conserved protein	NA	P → F
1414	5	1.250	Mitochondrial ribosomal protein L36 [*Culex quinquefasciatus*]	NA	P → F
1513	3	∞	UBX domain-containing protein [*Pediculus humanus corporis*]	Zinc-ion binding	ns
1539	**2**	**-**	**ns**	**-**	**het -**
1555	1	-	Ribosomal biogenesis-like protein (*BRX1*) homolog [*Nasonia vitripennis*]	RNA binding	ns
1667	3	-	Similar to *Gryllus bimaculatus* mRNA, GB12028	NA	ns
1724	2	-	Recombination activating gene 1 activating-like protein (*Rag1ap1*) [*Nasonia vitripennis*]	Glucoside transmembrane transporter activity	het +
1774	5	0	ns	-	P → F
1790	3	-	ns	-	F → P
1851	2	-	ns	-	P → F
2100	3	0.641	Fructose-1,6-bisphosphatase-like protein [*Bombyx mori*]	Hydrolysis of fructose-1,6-bisphosphate	P → F
2182	3	-	Histone H2A-like protein [*Ixodes scapularis*]	DNA binding	ns
2271	1	-	ns	-	F → P
2361	1	-	ns	-	ns
2467	**1**	**-**	**Talin-2-like protein [*Pediculus humanus corporis*]**	**Actin binding; protein binding; structural molecular activity**	**het -**
2570	3	-	Similar to *Gryllus bimaculatus* mRNA, GB26520	NA	F → P
2733	7	-	Lipoic acid synthetase-like protein [*Pediculus humanus corporis*]	Metal ion binding	ns
2831	4	-	Similar to *Anopheles gambiae* AGAP002630 protein	NA	ns

Table S4.5 (Continued)

2833	2	∞	Aspartyl-tRNA synthetase [*Nasonia vitripennis*]	Nucleic acid binding	ns
2864	**4**	**∞**	**Mesencephalic astrocyte-derived neurotrophic factor (Manf) homolog [*Apis mellifera*]**	**Growth factor activity**	**het -**
2989	4	-	ns	-	ns
3084	3	-	Mitotic checkpoint MAD1-like protein [*Acyrthosiphon pisum*]	Protein binding	ns
3136	1	∞	Similar to *Pediculus humanus corporis* protein, PHUM494000	NA	P → F
3182	2	-	ns	-	F → P
3268	1	∞	Werner Syndrome-like exonnuclease (*WRNexo*)-like protein [*Bombus impatiens*]	Protein binding; nucleic acid binding; 3'-5' exonnuclease activity	ns
3344	3	∞	Arp2/3 complex 20 kD subunit [*Pediculus humanus corporis*]	Actin binding	ns
3422	**1**	**-**	**Microtubule-associated protein RP/EB family member 3 (*MAPRE3*)-like protein [*Tribolium castaneum*]**	**Microtubule binding; protein binding**	**het -**
3432	**3**	**∞**	**Similar to *DnaJ* chaperone (*Hsp40*) [*Aedes aegypti*]**	**Heat shock binding protein; ATP binding; protein binding; zinc ion binding**	**het -**
3528	1	-	Similar to *Tribolium castaneum* predicted protein CG11417	NA	het -
3555	1	∞	*ESF1* homolog [*Bombus impatiens*]	NA	ns
3838	**2**	**-**	**Putative growth hormone-inducible soluble protein [*Ixodes scapularis*]**	**Growth factor activity**	**het -**
3843	4	0	Eukaryotic translation initiation factor 4 (*eIF4*) [*Pediculus humanus corporis*]	Binding	het +
3968	2	0	Canopy-1 (*CNPY1*)-like protein [*Bombus terrestris*]	fibroblast growth factor activity; molecular activity	ns
4205	**1**	**-**	**Small G protein signaling modulator 3-like protein (*SGSM3*) [*Tribolium castaneum*]**	**Rab GTPase binding**	**het -**
4361	2	-	Similar to *Drosophila melanogaster* mRNA, CG42308	NA	P → F

Table S4.5 (Continued)

4450	4	-	Similar to *Gryllus bimaculatus* mRNA, GB24459	NA	ns
4655	3	0.148	Similar to *Drosophila willistoni* protein, GK24843	NA	het +
4679	4	-	INO80 complex subunit B-like (zinc finger protein) [*Nasonia vitripennis*]	Metal ion binding; protein binding	het +
4913	**3**	**-**	**Similar to *Anopheles gambiae* mRNA (immunoglobulin domain protein), AGAP002737**	**Protein binding**	**het -**
5021	**3**	**-**	**ns**	**NA**	**het -**
5052	1	-	Similar to *Strongylocentrotus purpuratus* protein, MGC52920	NA	ns
5131	**3**	**-**	**Golgi-associated microtubule-binding protein (*HOOK3*) [*Culex quinquefasciatus*]**	**Microtubule binding**	**het -**
5136	**2**	**-**	**Similar to *Gryllus bimaculatus* protein GB05381**	**NA**	**het -**
5177	**1**	**-**	**ns**	**-**	**het -**
5214	**6**	**∞**	**ns**	**-**	**het -**
5368	5	0.411	ns	-	P → F
5556	2	0.244	Similar to *Gryllus bimaculatus* protein GB05381	NA	ns
5711	4	0.878	Polypeptide GalNAc transferase 6-like protein (*GALNT6*) [*Nasonia vitripennis*]	Protein glycosylation	F → P
5727	4	0.271	CSL-type zinc finger-like protein [*Pediculus humanus corporis*]	Molecular function (catalysis, binding); transferase activity	ns
5777	**3**	**∞**	**Similar to transport and Golgi organization 1 (*Tango1*) [*Apis mellifera*]**	**Molecular function (catalysis, binding)**	**het -**
5961	**1**	**∞**	**ns**	**-**	**het -**
6023	6	-	ns	-	ns
6026	3	-	Omega-amidase (*NIT2-B*) [*Taeniopygia guttata*]	Hydrolase activity	het +
6030	3	0.147	Ethanolaminephosphotransferase [*Anopheles gambiae*]	Transferase activity	ns
6057	2	-	Ubiquitin/ribosomal protein L40 [Bombyx mori]	Ribosome structural constituent	F → P
6128	**2**	**-**	**ns**	**-**	**het -**
6271	**3**	**-**	**ns**	**-**	**het -**
6557	3	-	Similar to *Gryllus bimaculatus* mRNA, GB07377	NA	ns

Table S4.5 (Continued)

6571	2	-	ns	-	ns
6579	1	-	ns	-	het +
6718	2	-	ns	-	P → F
6771	**1**	**∞**	**Dihydrolipoami de branched chain transacylase E2 (*DBT*) [*Xenopus laevis*]**	**Protein binding; cofactor binding; transferase activity**	**het -**
7046	6	0.184	ns	-	ns
7153	8	0.097	Proliferating cell nuclear antigen (*PCNA*) [*Aedes aegypti*]	DNA binding	P → F
7164	**4**	**-**	**ns**	**-**	**het -**
7469	1	0	Alpha-aspartyl dipeptidase-like protein [*Gallus gallus*]	Serine peptidase	ns
7566	1	-	ns	-	P → F
8026	2	0	Similar to *Anopheles gambiae* protein AGAP003185	Acyl CoA binding	ns
8229	**1**	**-**	**Cytochrome C oxidase subunit polypeptide 2-like protein [*Gallus gallus*]**	**Electron carrier activity**	**het -**
8257	1	-	Similar to *Gryllus bimaculatus* GB30455 protein	NA	P → F
8322	2	-	ns	-	F → P
8354	1	-	ns	-	ns
8375	**1**	**∞**	**Calsyntenin-1-like protein (*Clstn1*) [*Nasonia vitripennis*]**	**Calcium ion binding**	**het -**
8612	**2**	**0.350**	**Similar to *Helobdella robusta* clone CH306**	**NA**	**het -**
9839	2	-	ns	-	P → F
9851	3	0	ns	-	P → F
10368	**2**	**-**	**Similar to *Apis mellifera* protein CG14232**	**Binding, fatty-acy-CoA binding**	**het -**
11695	1	-	Vacuolar proton ATPase [*Aedes aegypti*]	Hydrogen ion transmembrane transporter activity	het +
12397	**1**	**-**	**ns**	**-**	**het -**
14713	**4**	**-**	**ns**	**-**	**het -**
14937	2	-	ns	-	ns
16015	1	-	ns	-	ns

Figure S4.1 Individual genomic clines for all 125 markers analyzed. Each genomic cline depicts the extent of introgression (calculated based on 301 individuals from within the Pennsylvania hybrid zone) for a focal locus. The hybrid index represents genome-wide admixture based on the proportion of alleles inherited from *G. firmus* (hybrid index: 0 = *G. pennsylvanicus*, 1 = *G. firmus*). The shaded gray areas depict the 95% confidence intervals for the probability of observing a homozygous *G. pennsylvanicus* genotype (dark grey) or heterozygous genotype (light grey) at the focal locus, given the hybrid index. Observed genotype classes are plotted against the hybrid index as open circles (top: homozygous *G. pennsylvanicus*, middle: heterozygote, bottom homozygous *G. firmus*), and the frequency of observed genotypes is indicated to the right of the panel. The genomic clines for homozygous *G. pennsylvanicus* genotype (solid line) and heterozygous genotype (dashed line) are overlaid onto the 95% confidence intervals for the genotype probabilities; genomics clines that fall outside of the expected distribution represent significant deviations from neutral expectations. Excess (+) or deficit (-) of observed genotype classes are indicated above.

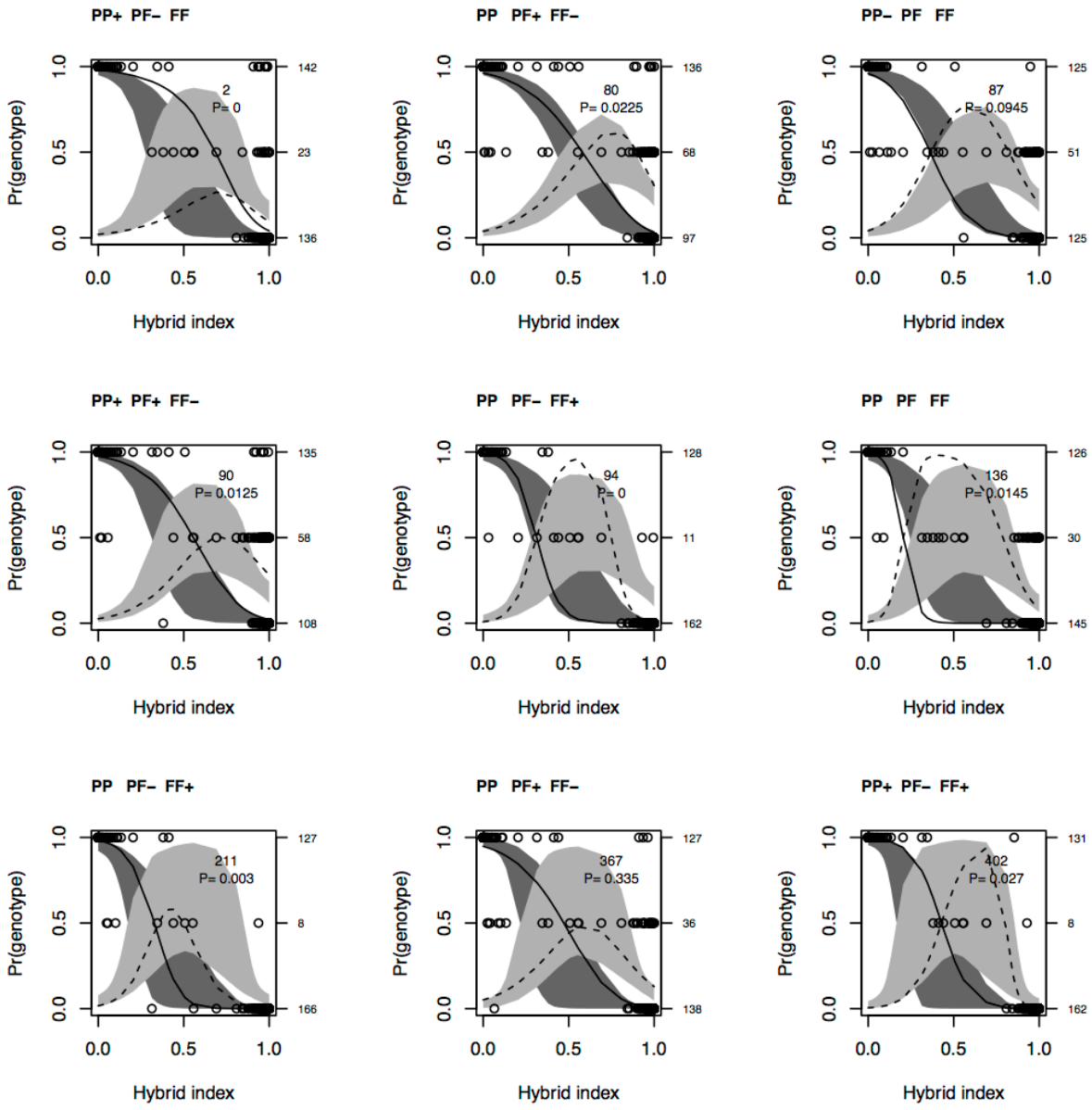

PP+ PF− FF
PP PF+ FF−
PP− PF FF
PP+ PF+ FF−
PP PF− FF+
PP PF FF
PP PF− FF+
PP PF+ FF−
PP+ PF− FF+
Pr(genotype)
Hybrid index
0.0
0.5
1.0
2
P= 0
80
P= 0.0225
87
P= 0.0945
90
P= 0.0125
94
P= 0
136
P= 0.0145
211
P= 0.003
367
P= 0.335
402
P= 0.027

Figure S4.1 (Continued)

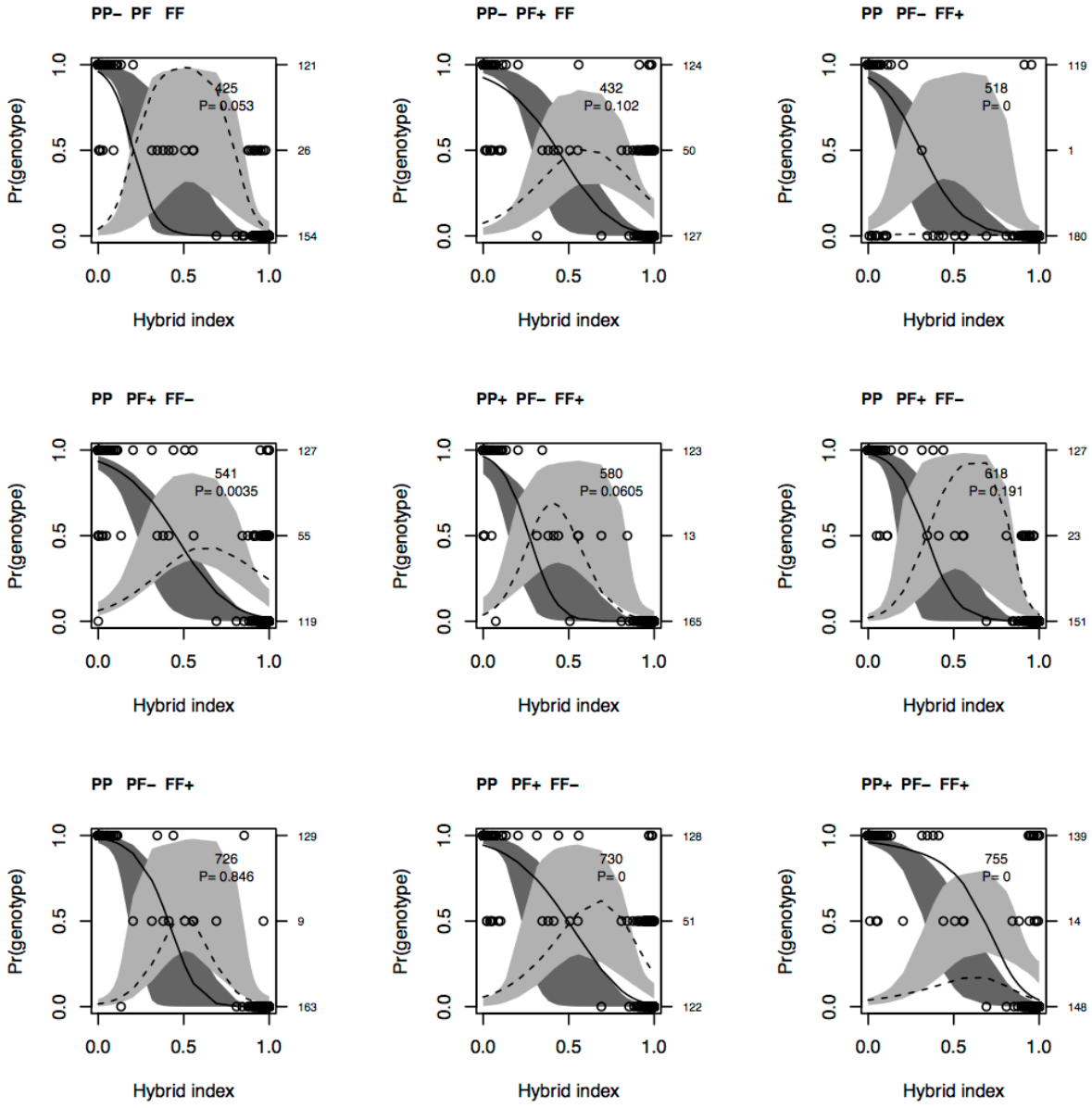


Figure S4.1 (Continued)

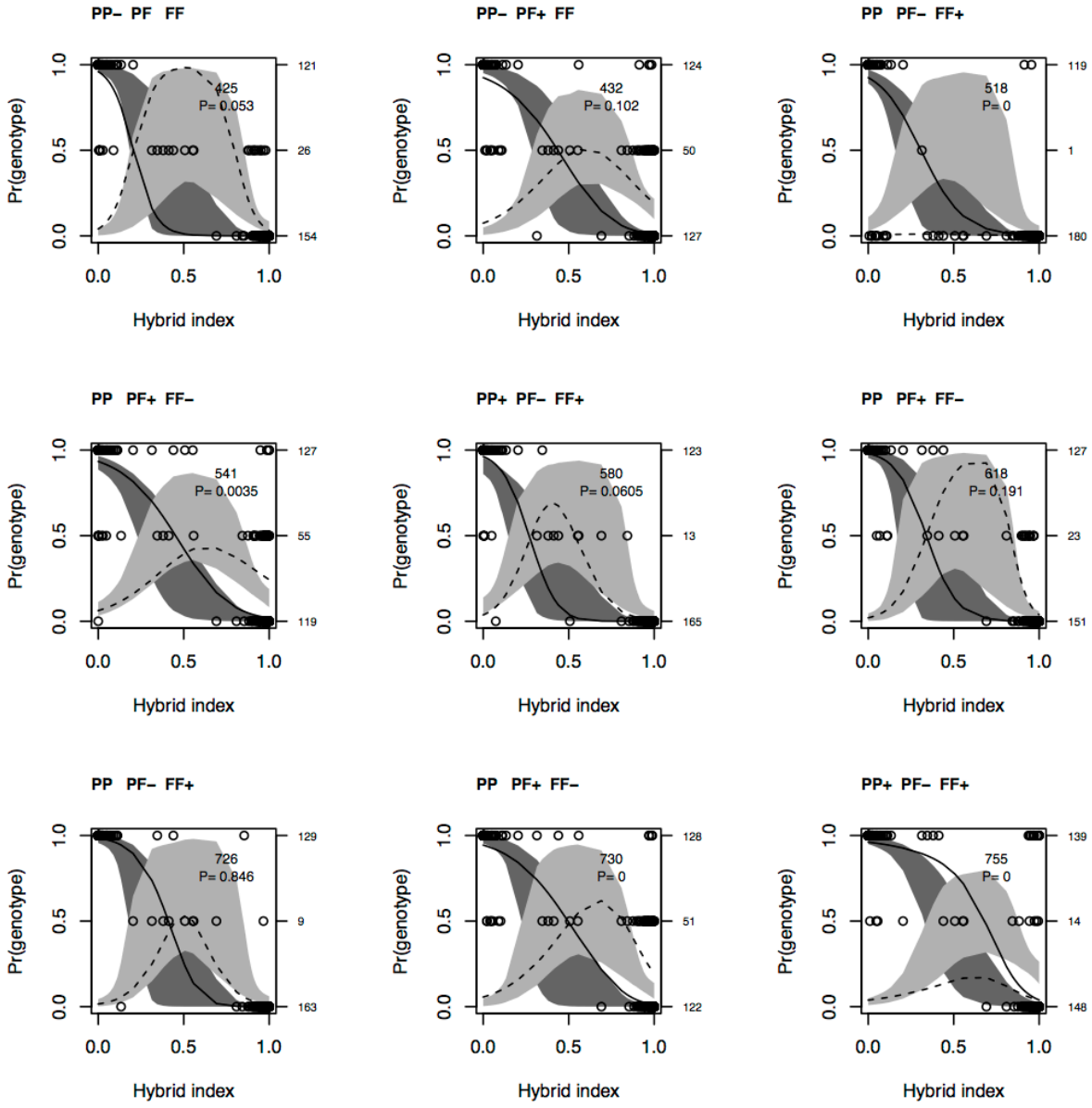


Figure S4.1 (Continued)

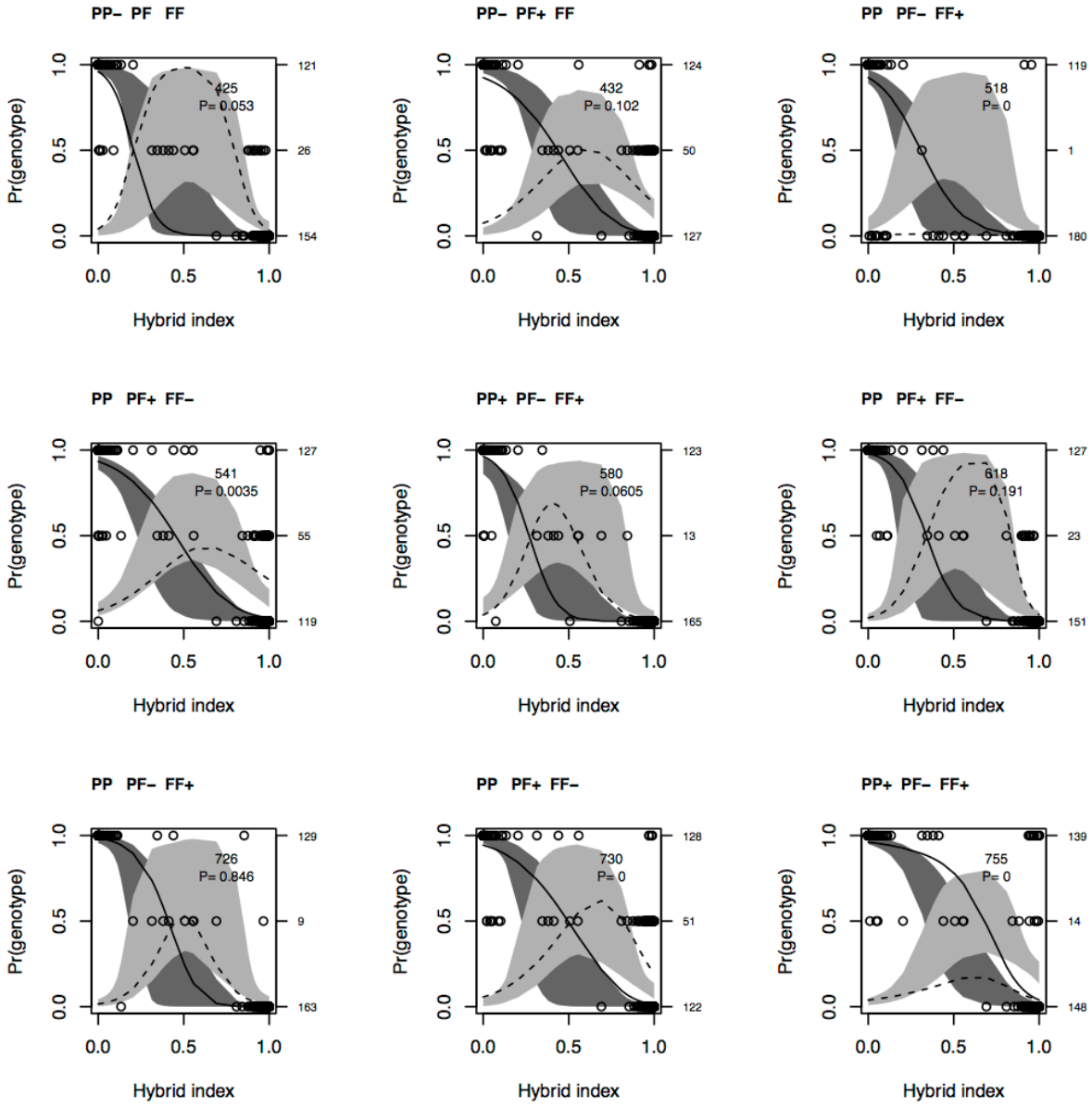


Figure S4.1 (Continued)

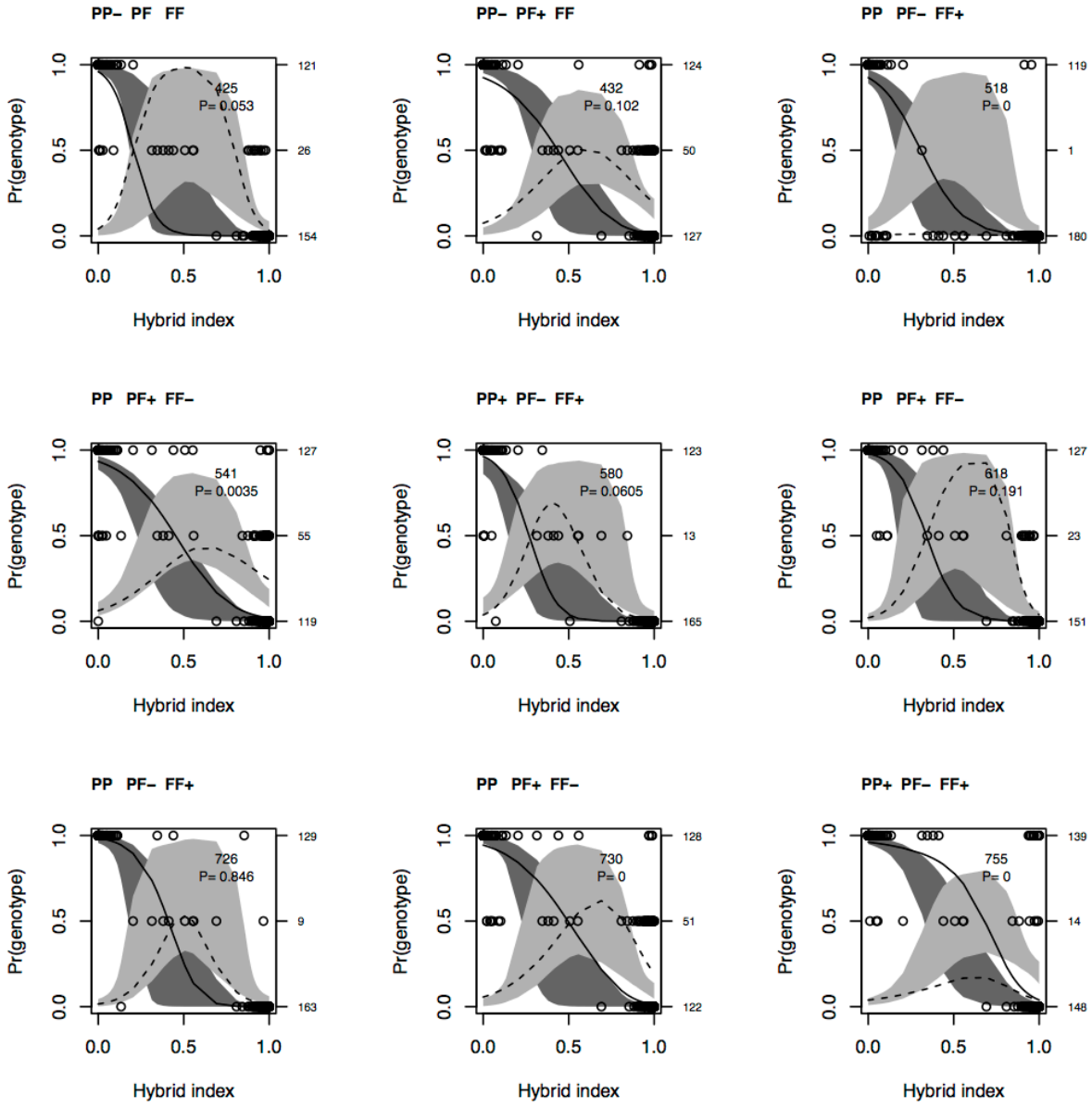


Figure S4.1 (Continued)

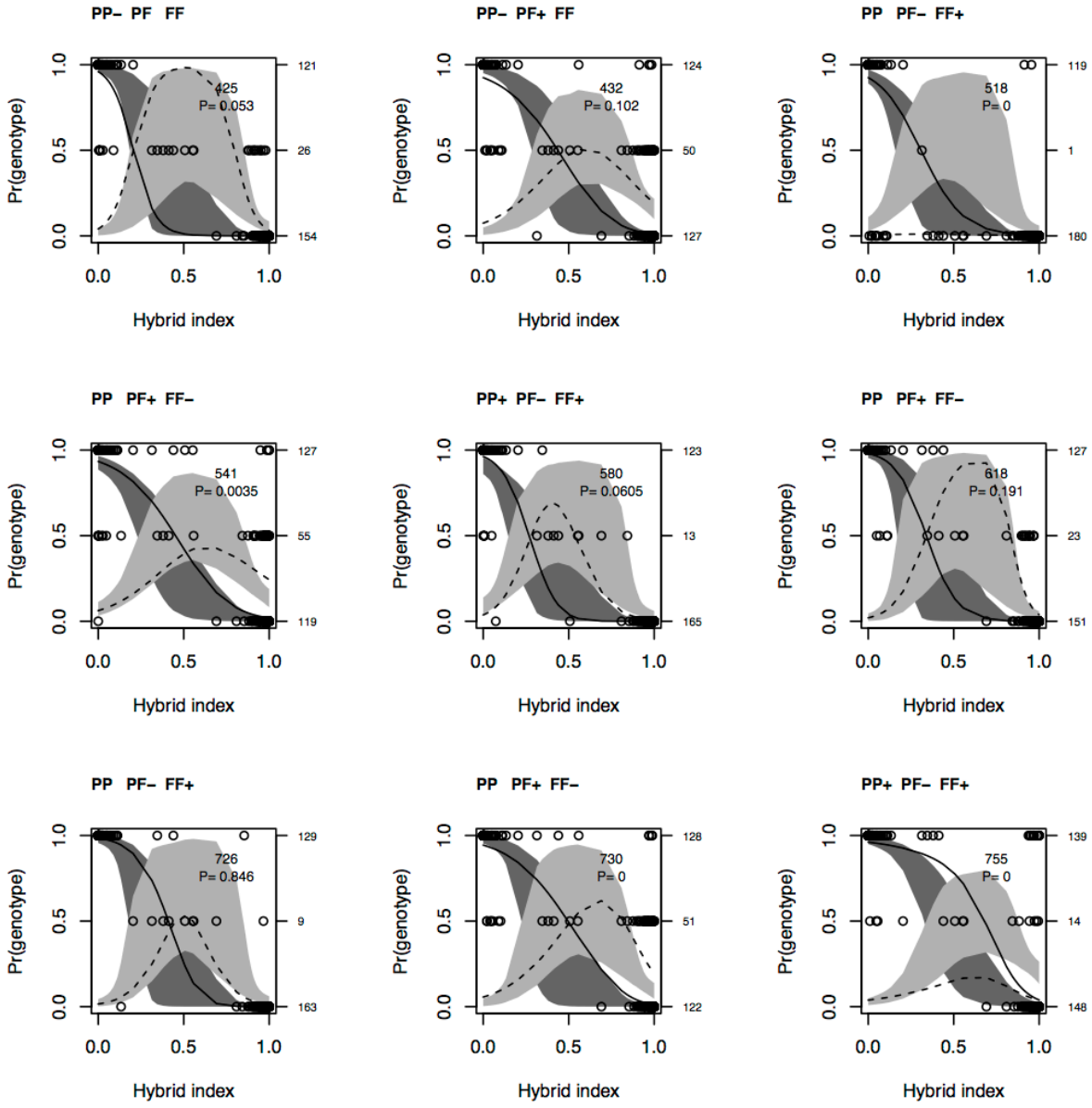


Figure S4.1 (Continued)

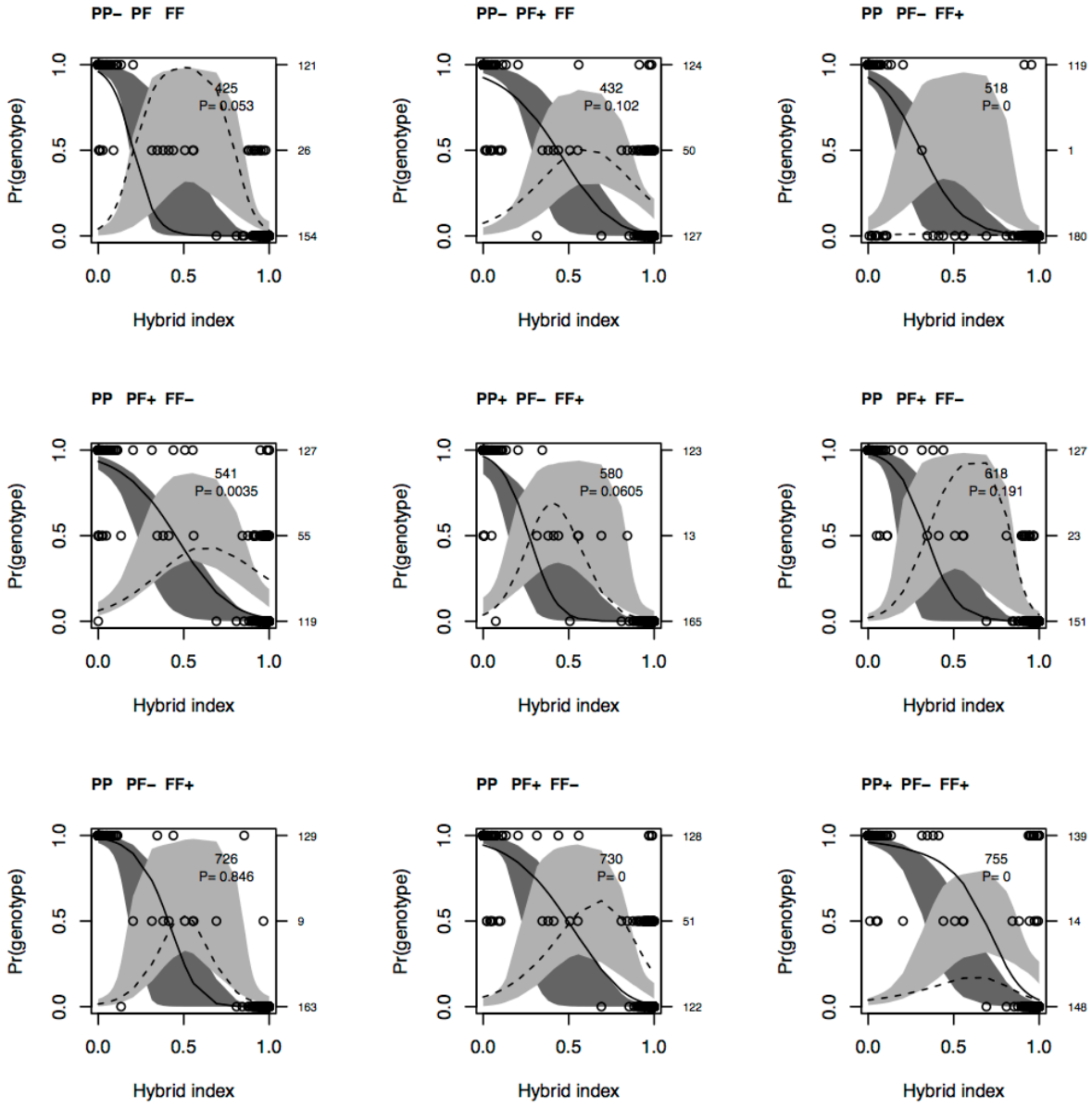


Figure S4.1 (Continued)

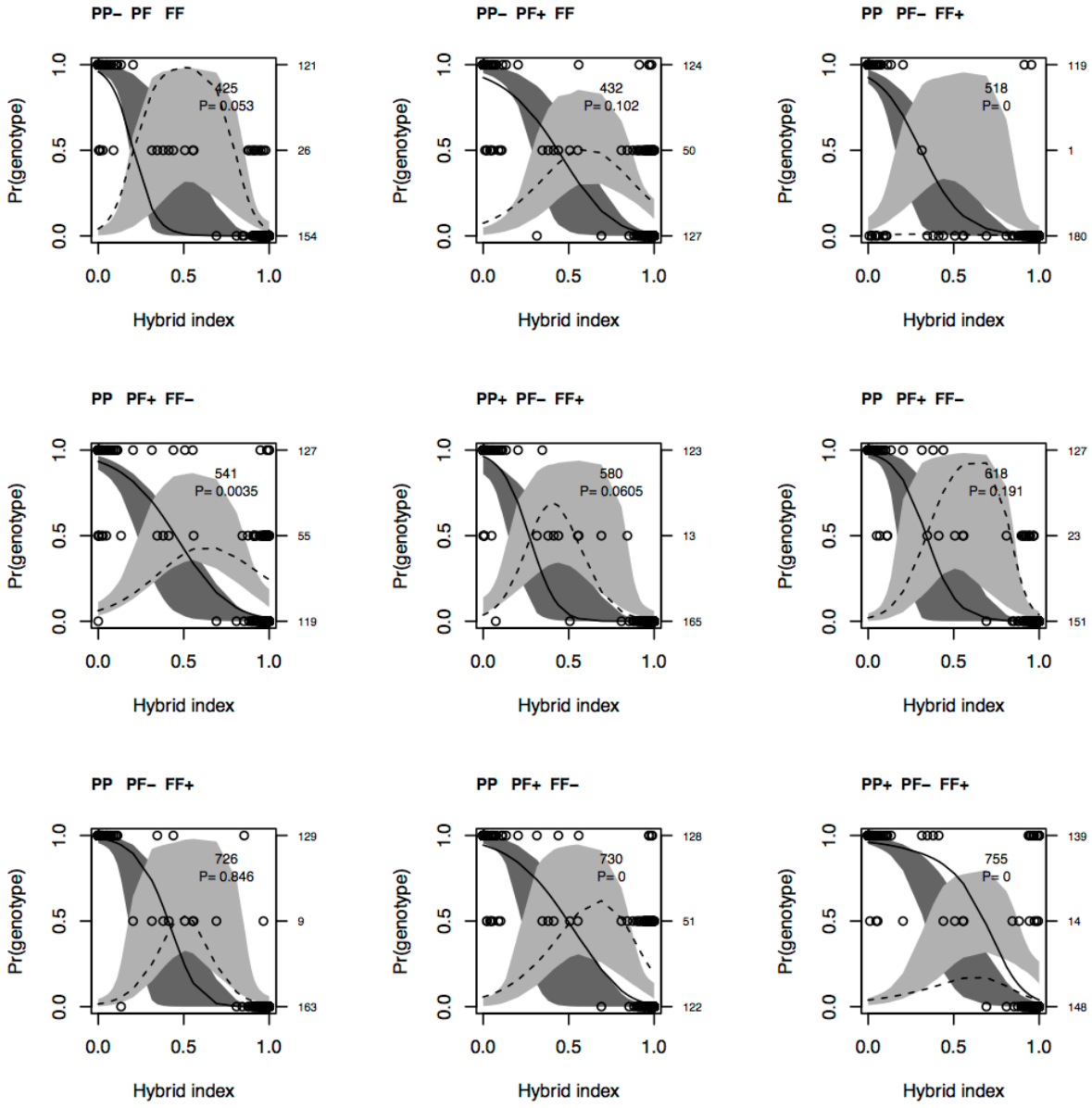


Figure S4.1 (Continued)

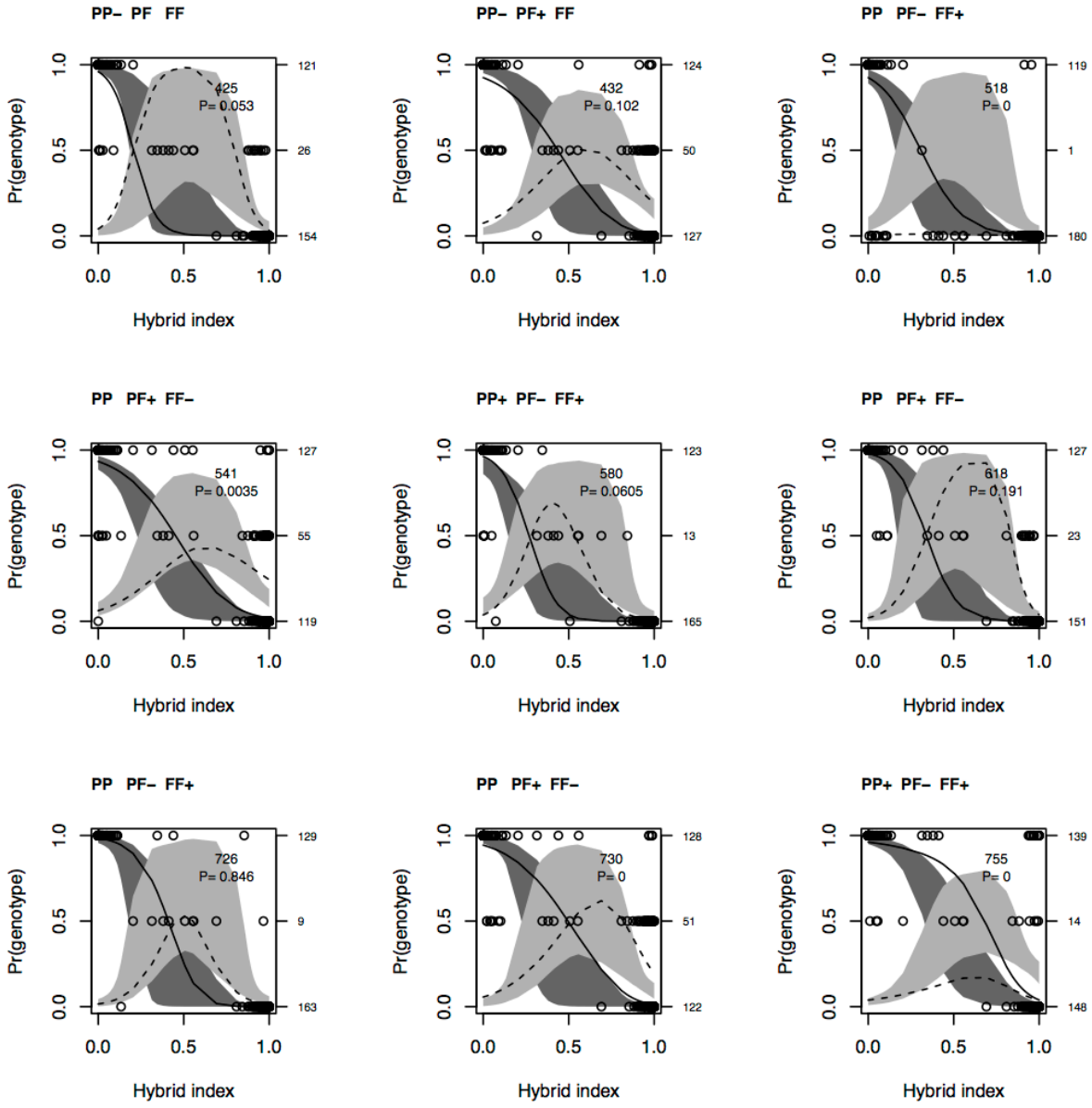


Figure S4.1 (Continued)

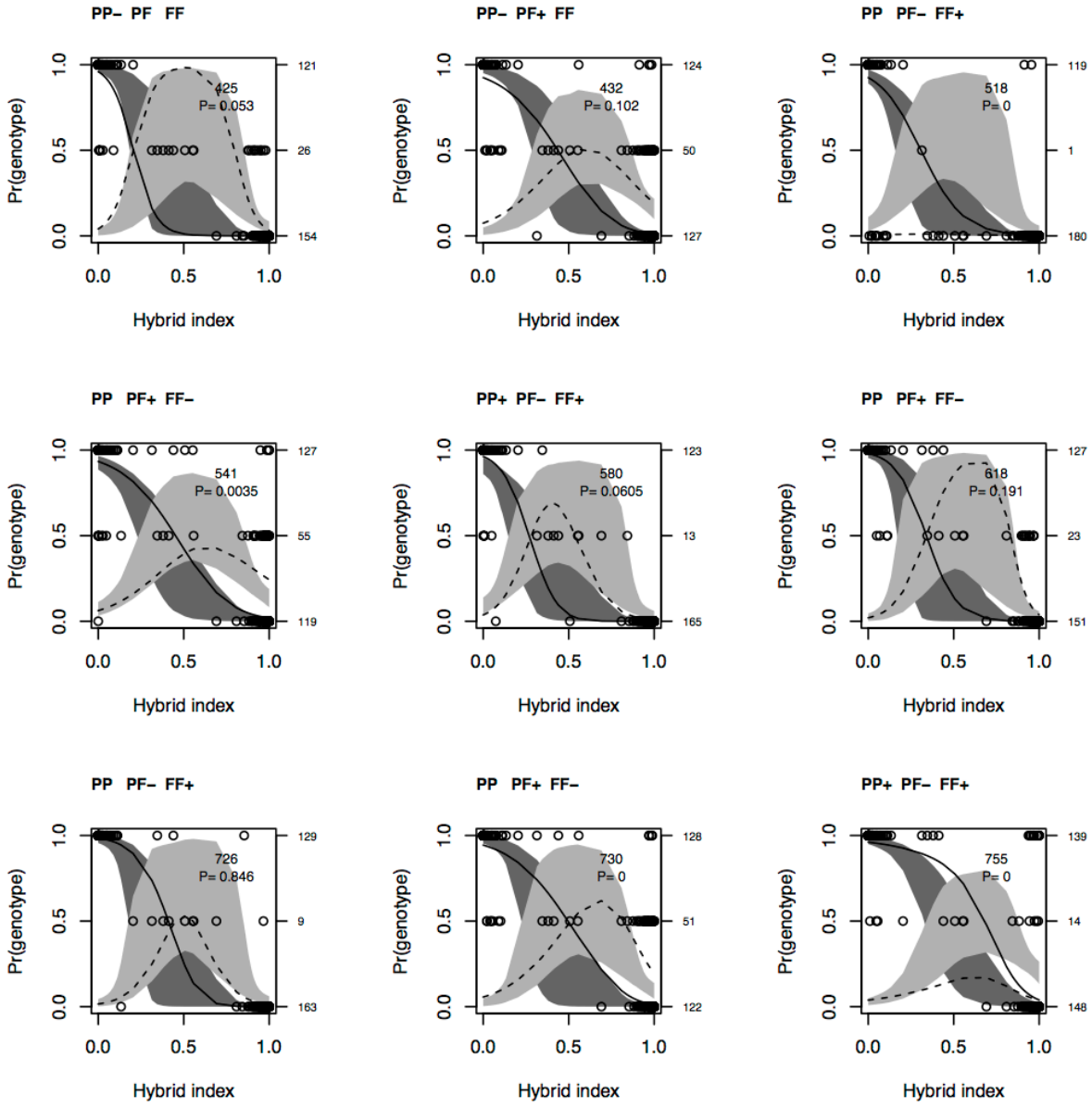


Figure S4.1 (Continued)

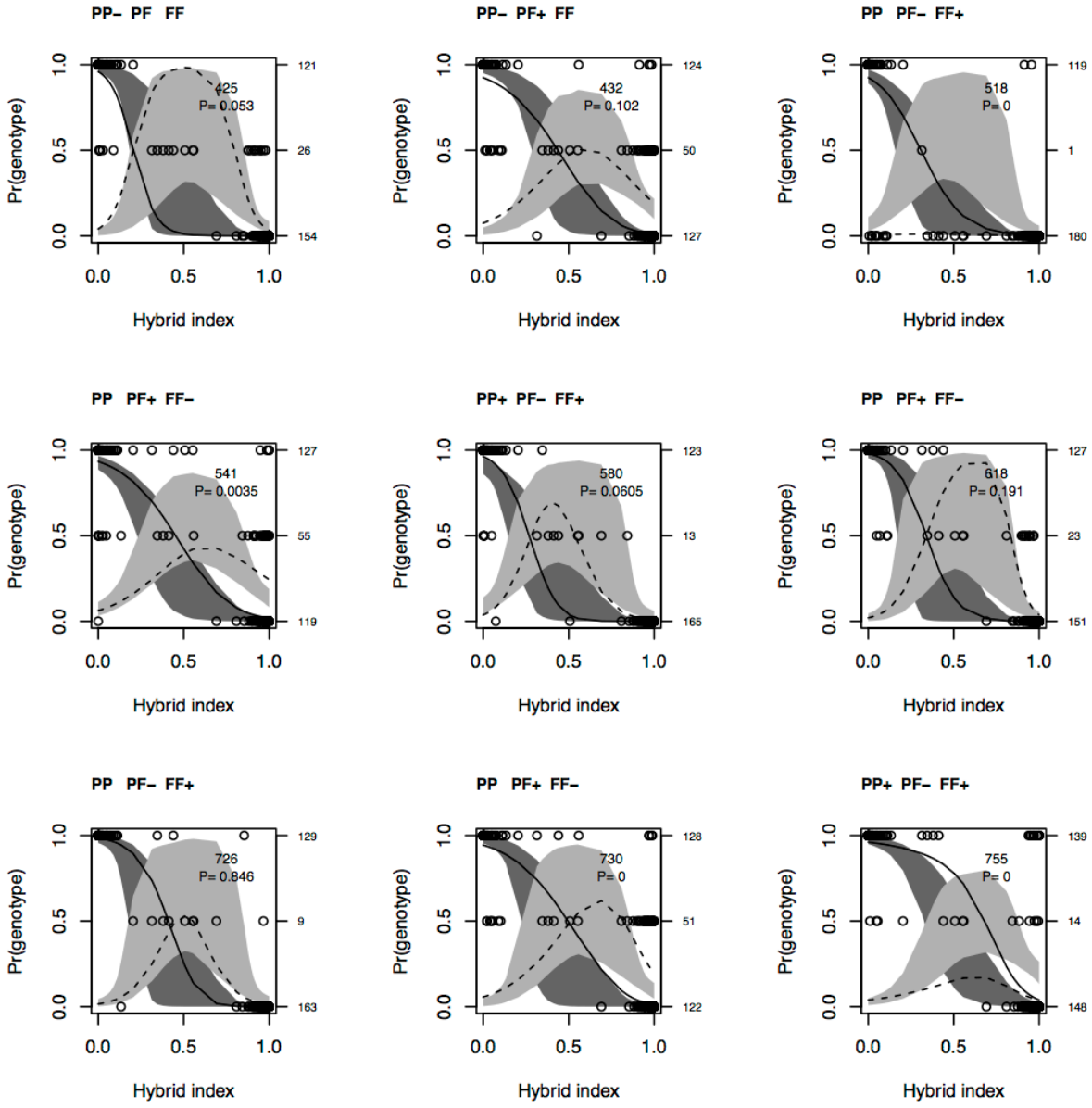


Figure S4.1 (Continued)

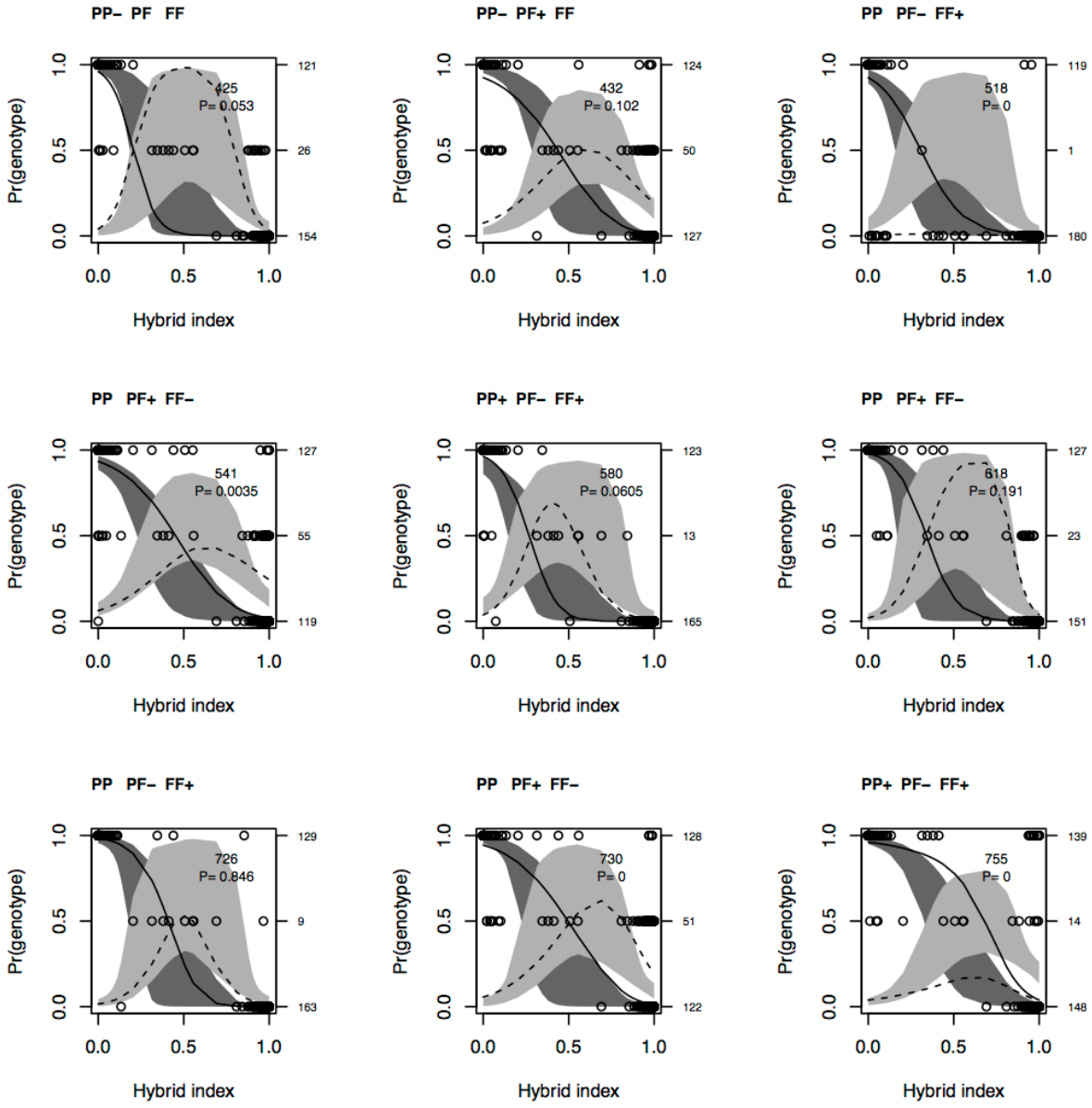
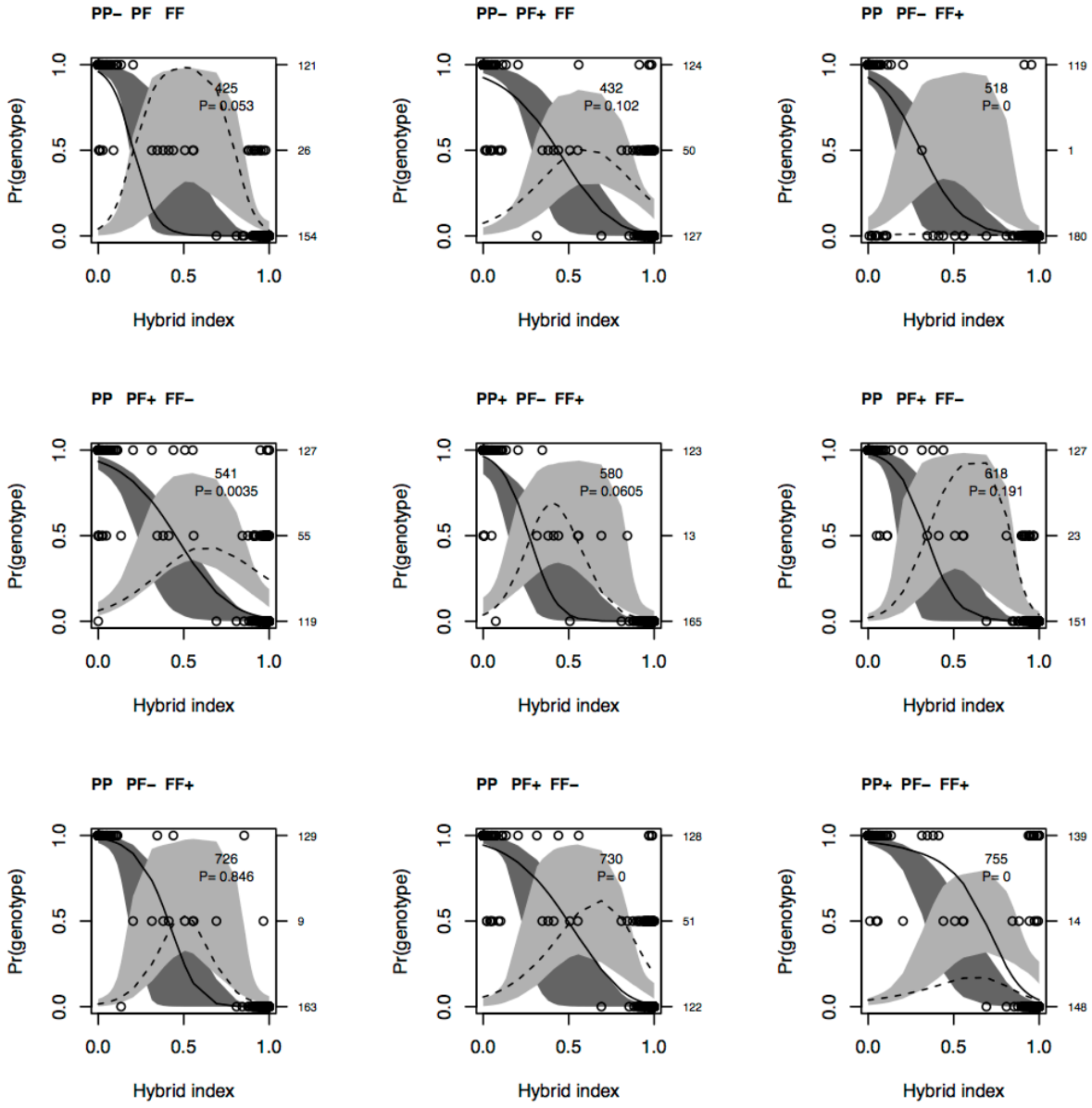


Figure S4.1 (Continued)



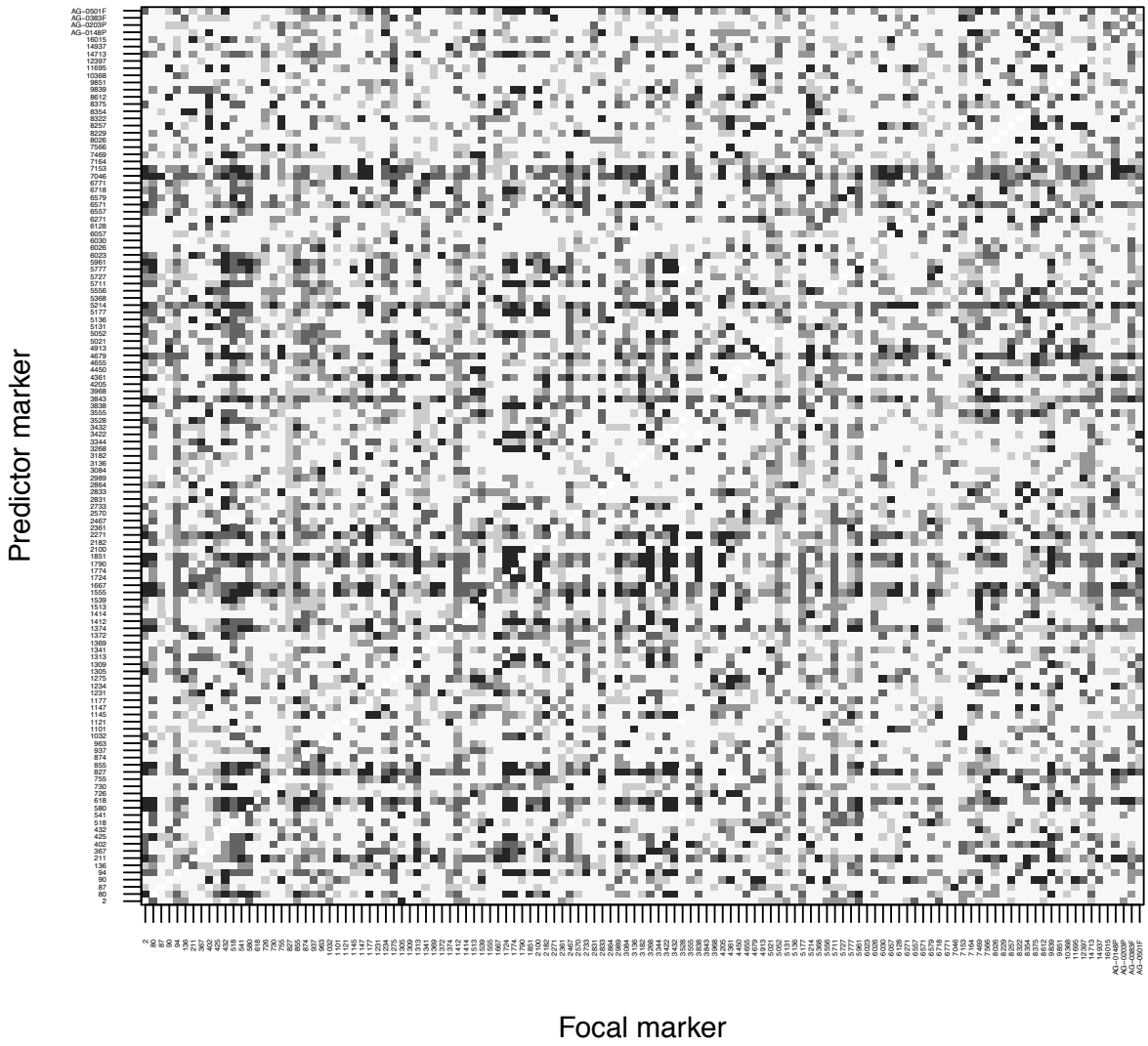


Figure S4.2 Plot of pairwise associations between 125 markers used in genomic cline analysis. Association were tested by using a regression model that includes a locus to predict genotypes at a focal locus (given the hybrid index and interspecific heterozygosity) and comparing the AIC values for this model with a regression model that does not include a predictor locus.
Boxes are shaded to denote the strength of association between loci, with darker grey denoting higher positive differences in AIC values between the two models.

Chapter 5

The following are supplementary tables and figures for Chapter 5.

Table S5.1 Results of genomic and geographic cline analysis for 114 markers genotyped in both the Pennsylvania and Connecticut regions of the hybrid zone between *G. firmus* and *G. pennsylvanicus*. In bold are genes that had significantly fewer heterozygotes than expected in both regions of the hybrid zone. [1] possible sex-linkage of marker; [2] likelihood ratio; [3] probability of departure from neutrality following false discovery rate correction (Benjamini and Hockber 1995); [4] over (+) or underrepresentation (-) of observed genotypes; [5] inferred deviation category; [6] geographic cline width and [7] geographic cline center.

		Pennsylvania							Connecticut							
Name	**[1]X**	**[2]LnL**	**[3]P**		**[4]Genotypes**			**[5]Dev.**	**[2]LnL**	**[3]P**		**[4]Genotypes**			**[5]Dev.**	**[6]*w***	**[7]*c***
2	?	15.79	0	*	*PP+*	*PF-*	*FF*	P → F	19.37	0	*	*PP*	*PF-*	*FF+*	*het -*	*346.1*	*311.3*
80		5.14	0.032	*	PP	PF+	FF-	het +	18.89	0	*	PP-	PF+	FF	het +	445.8	311.9
87		3.89	0.097		PP-	PF	FF	n.s.	7.68	0.006	*	PP-	PF+	FF	het +	411.6	322.7
90		7.14	0.018	*	PP+	PF+	FF-	P → F	52.52	0	*	PP+	PF+	FF-	P → F	603.4	429.2
94	**?**	**33.05**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**20.85**	**0**	*	**PP+**	**PF-**	**FF+**	**het -**	**229.9**	**329.4**
136		8.26	0.012	*	PP	PF	FF	P → F	15.12	0	*	PP+	PF	FF-	P → F	304.9	354.9
211		**11.42**	**0.002**	*	**PP**	**PF-**	**FF+**	**het -**	**11.19**	**0.002**	*	**PP+**	**PF-**	**FF+**	**het -**	**239.5**	**306.7**
367		1.30	0.487		PP	PF	FF	n.s.	24.59	0	*	PP	PF+	FF-	P → F	375.4	362.6
402		**7.76**	**0.015**	*	**PP+**	**PF-**	**FF+**	**het -**	**16.60**	**0**	*	**PP+**	**PF-**	**FF**	**het -**	**193.8**	**323.0**
425		5.69	0.057		PP-	PF	FF	n.s.	3.88	0.120		PP	PF	FF	n.s.	295.0	303.1
432		3.39	0.153		PP-	PF+	FF	n.s.	0.56	0.856		PP	PF	FF	n.s.	317.7	332.8
518	?	67.72	0	*	PP	PF-	FF+	F → P	157.93	0	*	PP-	PF-	FF+	F → P	449.0	188.9
541		8.33	0.008	*	PP	PF+	FF-	het +	35.09	0	*	PP-	PF+	FF-	het +	777.8	292.7
618		2.90	0.233		PP	PF	FF-	n.s.	8.79	0.002	*	PP+	PF	FF-	P → F	248.1	315.9
726		0.22	0.712		PP	PF-	FF+	n.s.	22.05	0	*	PP+	PF-	FF	het -	196.1	318.0
730		14.28	0	*	PP	PF+	FF-	P → F	10.67	0.001	*	PP	PF+	FF-	P → F	344.4	342.3
755		28.28	0	*	PP+	PF-	FF+	P → F	9.01	0.002	*	PP+	PF-	FF	P → F	331.2	333.5
827	?	1.45	0.462		PP	PF-	FF+	n.s.	18.66	0	*	PP+	PF-	FF+	het -	212.8	314.2

Table S5.1 (Continued)

855		21.14	0	*	PP+	PF+	FF-	P → F	50.94	0	*	PP	PF+	FF-	P → F	869.6	513.6
874	**?**	**12.23**	**0.002**	*	**PP**	**PF-**	**FF+**	**het -**	**15.27**	**0**	*	**PP**	**PF-**	**FF+**	**het -**		
937		45.74	0	*	*PP+*	*PF*	*FF-*	P → F	20.96	0	*	*PP-*	*PF+*	*FF*	F → P	502.7	283.9
963	**?**	**13.25**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**6.48**	**0.020**	*	**PP**	**PF-**	**FF+**	**het -**	**290.5**	**293.9**
1032		0.60	0.802		PP	PF	FF	n.s.	43.01	0	*	PP	PF+	FF-	het +	744.1	432.0
1121	?	17.58	0	*	PP	PF-	FF+	het -	18.67	0	*	PP	PF-	FF+	het -	274.7	305.1
1145		28.74	0	*	PP-	PF	FF+	F → P	78.72	0	*	PP-	PF+	FF+	F → P	693.9	125.3
1177		5.95	0.044		PP	PF+	FF-	n.s.	3.06	0.192		PP	PF	FF	n.s.	275.5	298.9
1231		2.86	0.243		PP	PF	FF-	n.s.	16.31	0	*	PP+	PF	FF-	P → F	265.0	325.4
1234		2.04	0.332		PP	PF	FF	n.s.	10.26	0	*	PP	PF	FF	P → F	315.0	346.1
1275		12.54	0.002	*	PP+	PF+	FF-	P → F	8.13	0.002	*	PP+	PF	FF	P → F	279.4	343.6
1309		12.31	0.002	*	PP	PF+	FF-	het +	51.99	0	*	PP-	PF+	FF-	het +	582.2	375.9
1313		9.95	0.004	*	PP	PF+	FF-	P → F	13.18	0	*	PP+	PF	FF-	P → F	293.6	337.0
1341		4.34	0.120		PP	PF+	FF-	n.s.	33.72	0	*	PP	PF-	FF+	F → P	246.3	281.6
1369		62.12	0	*	PP	PF-	FF+	F → P	149.71	0	*	PP-	PF-	FF+	F → P	674.3	184.8
1372	?	15.25	0	*	PP+	PF-	FF	het -	30.87	0	*	PP+	PF-	FF	het -	204.9	327.8
1374	**?**	**7.27**	**0.027**	*	**PP**	**PF-**	**FF+**	**het -**	**22.05**	**0**	*	**PP+**	**PF-**	**FF**	**het -**	**273.7**	**338.8**
1412		8.12	0.015	*	PP+	PF	FF-	P → F	22.71	0	*	PP	PF	FF-	P → F	381.5	319.7
1513		4.37	0.107		PP+	PF-	FF	n.s.	14.03	0	*	PP	PF-	FF+	het -	266.0	292.1
1539		**14.88**	**0**	*	**PP+**	**PF-**	**FF+**	**het -**	**14.30**	**0**	*	**PP+**	**PF-**	**FF+**	**het -**	**237.1**	**313.0**
1555		1.19	0.526		PP	PF	FF-	n.s.	7.95	0.006	*	PP+	PF-	FF	het -	228.0	320.3
1667	?	3.78	0.162		PP+	PF-	FF	n.s.	27.26	0	*	PP+	PF-	FF	het -	183.3	320.0
1724		33.19	0	*	PP	PF+	FF-	het +	32.47	0	*	PP-	PF+	FF-	het +	667.2	235.5
1774		58.29	0	*	*PP+*	*PF+*	*FF-*	P → F	27.71	0	*	*PP-*	*PF+*	*FF-*	*het +*	453.4	328.2
1790		12.54	0	*	PP-	PF	FF+	F → P	54.46	0	*	PP-	PF+	FF+	F → P	583.2	182.2
1851		31.26	0	*	PP+	PF+	FF-	P → F	17.00	0	*	PP	PF+	FF-	P → F	365.9	346.4
2100		7.80	0.015	*	PP	PF+	FF-	P → F	3.45	0.166		PP	PF	FF	n.s.	268.0	308.8
2182	?	5.92	0.049		PP+	PF-	FF	n.s.	25.51	0	*	PP+	PF-	FF	het -	196.0	323.1
2271		21.06	0	*	PP-	PF	FF+	F → P	51.10	0	*	PP-	PF-	FF+	F → P	408.7	264.9

Table S5.1 (Continued)

2361		0.37	0.764		PP	PF	FF	n.s.	28.56	0	*	PP+	PF+	FF-	P → F	372.7	362.3
2467	**?**	**11.85**	**0**	*****	**PP+**	**PF-**	**FF**	**het -**	**19.91**	**0**	*****	**PP+**	**PF-**	**FF**	**het -**	**185.0**	**317.6**
2733		3.76	0.157		PP	PF	FF	n.s.	24.71	0	*	PP+	PF-	FF	het -	205.1	329.8
2831		2.71	0.229		PP	PF	FF	n.s.	10.39	0.001	*	PP	PF-	FF+	F → P	274.3	306.4
2833	?	0.16	0.728		PP	PF-	FF+	n.s.	19.96	0	*	PP+	PF-	FF	het -	185.7	316.9
2864	**?**	**9.80**	**0.004**	*****	**PP+**	**PF-**	**FF+**	**het -**	**28.80**	**0**	*****	**PP+**	**PF-**	**FF+**	**het -**	**191.1**	**315.3**
2989		2.72	0.273		PP+	PF-	FF+	n.s.	9.66	0.001	*	PP+	PF-	FF	het -	242.6	320.8
3084		1.28	0.495		PP-	PF	FF	n.s.	8.75	0.002	*	PP	PF+	FF-	het +	421.0	353.4
3136	?	14.30	0.001	*	PP	PF-	FF+	P → F	21.99	0	*	PP	PF-	FF+	P → F	366.9	324.4
3182		12.94	0	*	PP-	PF	FF+	F → P	25.02	0	*	PP	PF-	FF+	F → P	316.1	289.3
3268		4.77	0.093		PP-	PF+	FF-	n.s.	37.81	0	*	PP-	PF+	FF	het +	523.8	259.2
3344	?	5.91	0.052		PP+	PF-	FF	n.s.	24.41	0	*	PP+	PF-	FF	het -	191.1	315.3
3422	**?**	**7.90**	**0.019**	*****	**PP**	**PF-**	**FF+**	**het -**	**28.86**	**0**	*****	**PP+**	**PF-**	**FF+**	**het -**	**191.1**	**315.3**
3432		**10.06**	**0.005**	*****	**PP+**	**PF-**	**FF+**	**het -**	**13.14**	**0**	*****	**PP+**	**PF**	**FF-**	**het -**	**302.8**	**332.6**
3528		15.12	0	*	PP	PF-	FF+	het -	7.37	0.008	*	PP+	PF-	FF	het -	253.3	314.7
3555	?	0.24	0.711		PP	PF-	FF+	n.s.	22.19	0	*	PP+	PF-	FF	het -	196.1	318.0
3838	?	13.29	0	*	**PP**	**PF-**	**FF+**	**het -**	6.51	0.024	*	**PP**	**PF-**	**FF+**	**het -**	**290.5**	**293.9**
3843		6.35	0.038		PP	PF+	FF-	n.s.	18.10	0	*	PP-	PF+	FF	het +	430.1	274.5
3968	?	3.84	0.143		PP+	PF-	FF	n.s.	27.13	0	*	PP+	PF-	FF	het -	183.3	320.0
4205		**10.55**	**0.004**	*****	**PP**	**PF-**	**FF+**	**het -**	**12.71**	**0**	*****	**PP**	**PF-**	**FF+**	**het -**	**277.0**	**292.4**
4361	?	9.87	0.004	*	PP+	PF	FF-	P → F	21.83	0	*	PP+	PF	FF	P → F	374.6	315.3
4450		2.04	0.338		PP+	PF-	FF+	n.s.	18.59	0	*	PP+	PF-	FF	het -	215.4	319.2
4655		25.05	0	*	PP-	PF+	FF-	het +	143.92	0	*	PP-	PF+	FF+	het +	817.5	143.5
4679		11.89	0.003	*	PP	PF+	FF-	het +	50.07	0	*	PP-	PF+	FF-	het +	567.5	373.0
4913		**24.66**	**0**	*****	**PP**	**PF-**	**FF+**	**het -**	**8.45**	**0.002**	*****	**PP+**	**PF-**	**FF+**	**het -**	**232.2**	**304.0**
5021	**?**	**12.17**	**0.001**	*****	**PP**	**PF-**	**FF+**	**het -**	**15.28**	**0**	*****	**PP**	**PF-**	**FF+**	**het -**	**270.7**	**289.7**
5052		5.35	0.078		PP+	PF-	FF	n.s.	30.07	0	*	PP+	PF-	FF-	P → F	211.7	337.8
5131	**?**	**9.82**	**0.004**	*****	**PP**	**PF-**	**FF+**	**het -**	**22.03**	**0**	*****	**PP+**	**PF-**	**FF**	**het -**	**204.5**	**318.3**
5136	**?**	**10.18**	**0.003**	*****	**PP**	**PF-**	**FF+**	**het -**	**16.57**	**0**	*****	**PP**	**PF-**	**FF+**	**het -**	**268.7**	**292.2**

Table S5.1 (Continued)

5177		**24.33**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**17.85**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**282.6**	**297.6**
5214		**11.99**	**0**	*	**PP+**	**PF-**	**FF+**	**het -**	**17.42**	**0**	*	**PP+**	**PF-**	**FF**	**het -**	**190.4**	**320.3**
5368		39.66	0	*	*PP+*	*PF+*	*FF-*	P → F	20.16	0	*	*PP-*	*PF+*	*FF-*	*Het +*	453.4	328.2
5556		2.76	0.217		PP-	PF+	FF	n.s.	3.53	0.175		PP	PF+	FF	n.s.	365.6	331.2
5711		12.40	0.001	*	PP-	PF	FF+	F → P	37.44	0	*	PP-	PF+	FF+	F → P	543.0	206.7
5727		4.40	0.095		PP-	PF	FF	n.s.	8.17	0.005	*	PP-	PF+	FF	Het +	409.9	286.4
5777	**?**	**14.70**	**0**	*	**PP**	**PF**	**FF+**	**het -**	**15.36**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**271.0**	**288.6**
5961		**31.47**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**14.09**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**315.5**	**289.3**
6023		1.60	0.419		PP	PF+	FF-	n.s.	2.01	0.425		PP	PF	FF	n.s.	233.9	312.9
6026		10.12	0.001	*	PP	PF+	FF-	het +	48.26	0	*	PP-	PF+	FF	het +	828.1	248.8
6030	?	6.35	0.034		PP+	PF-	FF+	n.s.	24.58	0	*	PP+	PF-	FF	het -	170.5	327.2
6271		12.99	0.002	*	PP	PF-	FF+	het -	11.19	0	*	PP	PF-	FF+	het -	321.7	317.4
6557		-0.37	0.809		PP	PF	FF	n.s.	20.20	0	*	PP-	PF+	FF	het +	432.1	287.1
6571		4.21	0.122		PP	PF	FF	n.s.	15.08	0	*	PP+	PF-	FF	het -	238.5	318.1
6579		6.14	0.033	*	*PP*	*PF*	*FF*	*het +*	88.52	0	*	*PP-*	*PF+*	*FF+*	F → P	*449.0*	*205.9*
6718	?	30.15	0	*	*PP+*	*PF*	*FF-*	P → F	10.92	0.001	*	*PP+*	*PF-*	*FF*	*het -*	235.6	316.8
6771		**17.74**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**13.37**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**282.8**	**307.9**
7046		2.52	0.304		PP	PF	FF-	n.s.	50.50	0	*	PP-	PF+	FF+	F → P	733.8	173.2
7153		36.30	0	*	*PP+*	*PF+*	*FF-*	P → F	120.16	0	*	*PP-*	*PF+*	*FF+*	F → P	428.1	209.1
7164	**?**	**10.27**	**0.004**	*	**PP**	**PF-**	**FF+**	**het -**	**20.61**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**270.7**	**289.7**
7469		1.73	0.411		PP	PF	FF	n.s.	18.16	0	*	PP-	PF+	FF-	het +	584.2	361.5
7566		6.17	0.019	*	*PP+*	*PF*	*FF*	P → F	82.76	0	*	*PP-*	*PF+*	*FF-*	*het +*	828.3	362.3
8026		**7.18**	**0.019**	*	**PP**	**PF-**	**FF+**	**het -**	**6.16**	**0.021**	*	**PP+**	**PF**	**FF**	**het -**	**259.7**	**314.9**
8229	**?**	**15.92**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**12.57**	**0.002**	*	**PP**	**PF-**	**FF+**	**het -**	**263.4**	**289.5**
8257		8.03	0.013	*	PP	PF+	FF-	P → F	7.89	0.006	*	PP+	PF-	FF	P → F	228.0	320.3
8322		18.08	0	*	PP	PF-	FF+	F → P	0.72	0.843		PP	PF	FF	n.s.	246.7	296.8
8375		**10.82**	**0.004**	*	**PP**	**PF-**	**FF+**	**het -**	**13.15**	**0.001**	*	**PP**	**PF-**	**FF+**	**het -**	**277.0**	**292.4**
8612		**12.38**	**0.001**	*	**PP**	**PF-**	**FF+**	**het -**	**28.74**	**0**	*	**PP+**	**PF-**	**FF+**	**het -**	**285.1**	**295.1**
9839		23.12	0	*	PP+	PF-	FF	P → F	21.94	0	*	PP+	PF-	FF	P → F	417.9	354.6

Table S5.1 (Continued)

9851		45.07	0	*	*PP*+	*PF*+	*FF*-	P → F	20.78	0	*	*PP*-	*PF*+	*FF*-	*het* +	456.3	325.5
10368		**9.81**	**0.003**	*	**PP**	**PF-**	**FF+**	**het -**	**9.47**	**0.001**	*	**PP+**	**PF-**	**FF**	**het -**	**234.0**	**307.8**
11695		16.92	0	*	*PP*-	*PF*+	*FF*-	*het* +	128.18	0	*	*PP*-	*PF*+	*FF*+	F → P	617.7	166.0
12397	**?**	**12.10**	**0.003**	*	**PP**	**PF-**	**FF+**	**het -**	**15.04**	**0**	*	**PP**	**PF-**	**FF+**	**het -**	**270.7**	**289.7**
14937		2.92	0.246		PP	PF	FF-	n.s.	7.72	0.010	*	PP+	PF	FF	het +	257.3	314.9
16015		2.64	0.284		PP	PF	FF	n.s.	11.19	0	*	PP-	PF+	FF+	F → P	525.3	257.8
AG-0148P		13.99	0	*	PP	PF-	FF+	P → F	12.45	0	*	PP+	PF	FF-	P → F	304.0	375.3
AG-0383F		5.86	0.052		PP-	PF+	FF-	n.s.	27.92	0	*	PP-	PF+	FF-	het +	467.7	317.5
AG-0501F	?	3.70	0.184		PP+	PF	FF	n.s.	15.17	0	*	PP	PF-	FF+	het -	350.0	275.5

Table S5.2 Results of geographic cline analyses for mtDNA and three anonymous nuclear markers from Ross and Harrison (2002) and 114 SNP markers. CI: confidence interval.

	Cline center (*c*)			Cline width (*w*)		
	Center	**CI -**	**CI +**	**Width**	**CI +**	**CI -**
mtDNA	588.6	527.6	649.6	374.4	537.4	272.6
pUC5	313.8	277.1	350.5	336.6	447.1	258.8
pUC279	399.6	343.2	456.1	442.7	682.1	307.1
pUC351	342.5	307.2	377.7	316.1	426.1	239.5
2	314.2	292.3	336.0	364.1	431.2	310.6
80	313.4	287.6	339.1	464.6	555.5	394.7
87	322.3	298.2	346.3	418.4	498.1	356.0
90	428.4	389.9	466.9	593.9	744.2	488.0
94	329.4	312.2	346.7	240.1	285.6	203.1
136	354.6	334.4	374.9	312.2	371.6	264.9
211	306.7	289.3	324.1	248.2	294.7	210.1
367	361.8	338.6	385.0	379.2	453.4	321.1
402	323.2	307.4	339.0	203.6	243.6	171.0
425	302.9	283.5	322.4	302.8	357.8	258.0
432	332.3	312.0	352.5	318.6	377.8	271.1
518	189.1	163.3	214.9	473.6	564.4	403.6
541	293.9	256.5	331.4	766.0	968.6	628.2
618	315.9	298.1	333.7	258.3	306.5	219.1
726	318.3	302.3	334.2	207.7	248.3	174.5
730	343.5	321.7	365.4	355.1	422.3	301.8
755	335.7	314.7	356.6	335.1	397.7	285.1
827	309.0	292.0	326.0	237.5	282.5	200.7
855	496.9	433.0	560.8	813.2	1087.8	641.8
874	289.8	270.9	308.7	289.8	342.4	246.7
937	283.2	256.2	310.2	507.8	608.9	430.9
963	294.0	274.4	313.6	308.9	364.7	263.4
1032	429.6	382.3	476.9	737.5	949.4	596.5
1121	304.9	286.4	323.4	277.4	328.3	235.8
1145	123.6	86.5	160.8	692.5	867.1	571.2
1177	300.1	281.1	319.1	292.0	345.2	248.6
1231	325.4	306.9	343.9	274.7	325.8	233.3
1234	345.8	325.2	366.5	323.7	384.7	275.0
1275	341.7	322.4	361.0	291.5	346.2	247.6
1309	378.9	344.1	413.6	605.2	749.2	502.2
1313	336.9	317.2	356.6	302.7	359.2	257.4
1341	283.0	265.3	300.7	259.0	306.9	219.5
1369	184.7	151.8	217.6	656.4	809.4	547.2
1372	331.5	314.8	348.2	224.4	267.5	189.4
1374	338.6	319.7	357.5	281.2	333.8	238.7

Table 5.2 (Continued)

1412	320.0	296.6	343.4	401.8	477.5	342.2
1513	292.3	273.5	311.0	285.2	337.2	242.7
1539	312.5	295.9	329.2	227.7	271.2	192.0
1555	320.4	303.3	337.6	238.6	283.8	201.8
1667	320.3	304.9	335.7	193.6	232.2	162.1
1724	236.0	202.8	269.2	682.3	844.5	567.5
1774	329.6	303.4	355.8	467.6	560.6	396.3
1790	181.3	151.1	211.6	587.5	714.5	494.2
1851	345.8	323.3	368.3	370.1	440.8	314.4
2100	308.6	290.3	326.8	270.9	320.9	230.1
2182	323.3	307.4	339.2	205.7	246.0	172.9
2271	268.1	244.7	291.4	413.1	488.8	353.3
2361	363.4	340.0	386.8	381.6	456.6	323.1
2467	317.8	302.4	333.3	195.2	234.1	163.5
2733	329.9	313.6	346.2	214.4	256.0	180.6
2831	306.2	287.7	324.7	277.1	328.1	235.6
2833	317.8	302.4	333.3	195.2	234.1	163.5
2864	315.6	299.9	331.3	203.0	243.0	170.3
2989	320.8	303.2	338.5	252.6	299.9	214.0
3084	353.4	328.0	378.8	434.9	521.7	368.0
3136	324.6	301.9	347.4	384.6	456.8	327.6
3182	289.1	269.0	309.3	323.4	381.6	276.1
3268	260.0	232.6	287.3	523.9	628.9	444.4
3344	315.6	299.9	331.3	203.0	243.0	170.3
3422	315.6	299.9	331.3	203.0	243.0	170.3
3432	333.2	313.3	353.1	309.2	366.6	263.0
3528	314.7	296.8	332.7	261.6	310.2	221.9
3555	318.3	302.3	334.2	207.7	248.3	174.5
3838	294.0	274.4	313.6	308.9	364.7	263.4
3843	276.8	252.6	300.9	433.6	514.4	370.2
3968	320.3	304.9	335.7	193.6	232.2	162.1
4205	292.5	273.3	311.6	295.8	349.4	252.0
4361	315.5	292.5	338.5	394.9	468.7	336.6
4450	317.6	300.9	334.3	227.9	271.4	192.3
4655	142.1	100.2	184.0	821.2	1056.6	666.1
4679	377.2	344.0	410.5	577.7	711.3	480.9
4913	304.1	286.9	321.3	242.8	288.5	205.3
5021	289.8	270.9	308.7	289.8	342.4	246.7
5052	337.8	321.3	354.4	220.7	263.3	186.2
5131	318.5	302.3	334.6	214.0	255.5	180.1
5136	292.3	273.5	311.1	287.6	339.9	244.8

Table 5.2 (Continued)

5177	297.7	278.4	317.1	301.0	355.6	256.5
5214	320.6	304.9	336.2	200.4	239.9	168.1
5368	327.7	301.8	353.6	461.4	552.5	391.3
5556	333.8	311.2	356.4	376.7	447.8	320.5
5711	205.8	177.2	234.4	554.4	669.1	468.8
5727	288.8	265.2	312.4	416.7	494.1	355.8
5777	289.8	270.9	308.7	289.8	342.4	246.7
5961	289.1	269.0	309.2	322.8	380.9	275.6
6023	312.5	294.1	330.8	273.1	323.6	232.0
6026	245.2	206.9	283.5	812.2	1036.5	662.5
6030	327.0	312.5	341.5	167.6	202.4	139.4
6271	317.3	297.7	336.9	304.9	360.9	259.7
6557	282.6	258.1	307.0	441.6	524.6	376.7
6571	318.2	300.7	335.7	248.8	295.5	210.7
6579	205.6	181.1	230.2	444.3	527.0	379.8
6718	316.8	299.5	334.2	245.8	292.0	208.1
6771	307.6	288.8	326.5	285.3	337.6	242.8
7046	172.4	136.1	208.7	730.0	916.0	601.7
7153	208.8	185.0	232.7	425.6	503.6	364.3
7164	289.8	270.9	308.7	289.8	342.4	246.7
7469	363.6	329.8	397.4	605.6	747.1	503.7
7566	365.1	319.5	410.7	842.3	1094.5	678.7
8026	314.9	296.7	333.1	267.8	317.4	227.4
8229	286.0	267.0	304.9	291.3	344.1	248.1
8257	320.4	303.3	337.6	238.6	283.8	201.8
8322	296.9	279.1	314.6	259.4	307.5	219.9
8375	292.5	273.3	311.6	295.8	349.4	252.0
8612	296.4	277.1	315.6	298.8	353.0	254.6
9839	352.0	328.2	375.8	398.5	476.2	337.9
9851	326.8	300.6	353.1	470.5	564.1	398.8
10368	307.9	290.7	325.1	242.8	288.6	205.4
11695	165.0	132.9	197.1	622.3	763.4	520.5
12397	289.8	270.9	308.7	289.8	342.4	246.7
14937	314.9	296.7	333.0	267.3	316.8	226.9
16015	257.0	229.6	284.4	524.1	629.1	444.6
AG.0148P	374.9	354.3	395.4	311.9	373.0	263.6
AG.0383F	318.5	292.1	344.9	477.7	572.5	405.1
AG.0501F	289.8	270.9	308.7	289.8	342.4	246.7

Figure S5.1. Individual genomic clines for all 114 markers analyzed in **A)** Pennsylvania (N = 301) and **B)** Connecticut (N = 260). Each genomic cline depicts the extent of genomic introgression for a focal locus. The hybrid index represents genome-wide admixture based on the proportion of alleles inherited from *G. firmus* (hybrid index: 0 = *G. pennsylvanicus*, 1 = *G. firmus*). The shaded gray areas depict the 95% confidence intervals for the probability of observing a homozygous *G. pennsylvanicus* genotype (dark grey) or heterozygous genotype (light grey) at the focal locus given the hybrid index. Observed genotype classes are plotted against the hybrid index as open circles (top: homozygous *G. pennsylvanicus*, middle: heterozygous, bottom homozygous *G. firmus*), and the frequency of observed genotypes is indicated on the right of the panel. The genomic clines for homozygous *G. pennsylvanicus* genotype (solid line) and heterozygous genotype (dashed line) are overlaid onto the 95% confidence intervals for the genotype probabilities; genomics clines that fall outside of the expected distribution represent significant deviations from neutral expectations. Excess (+) or deficit (-) of observed genotype classes are indicated above.

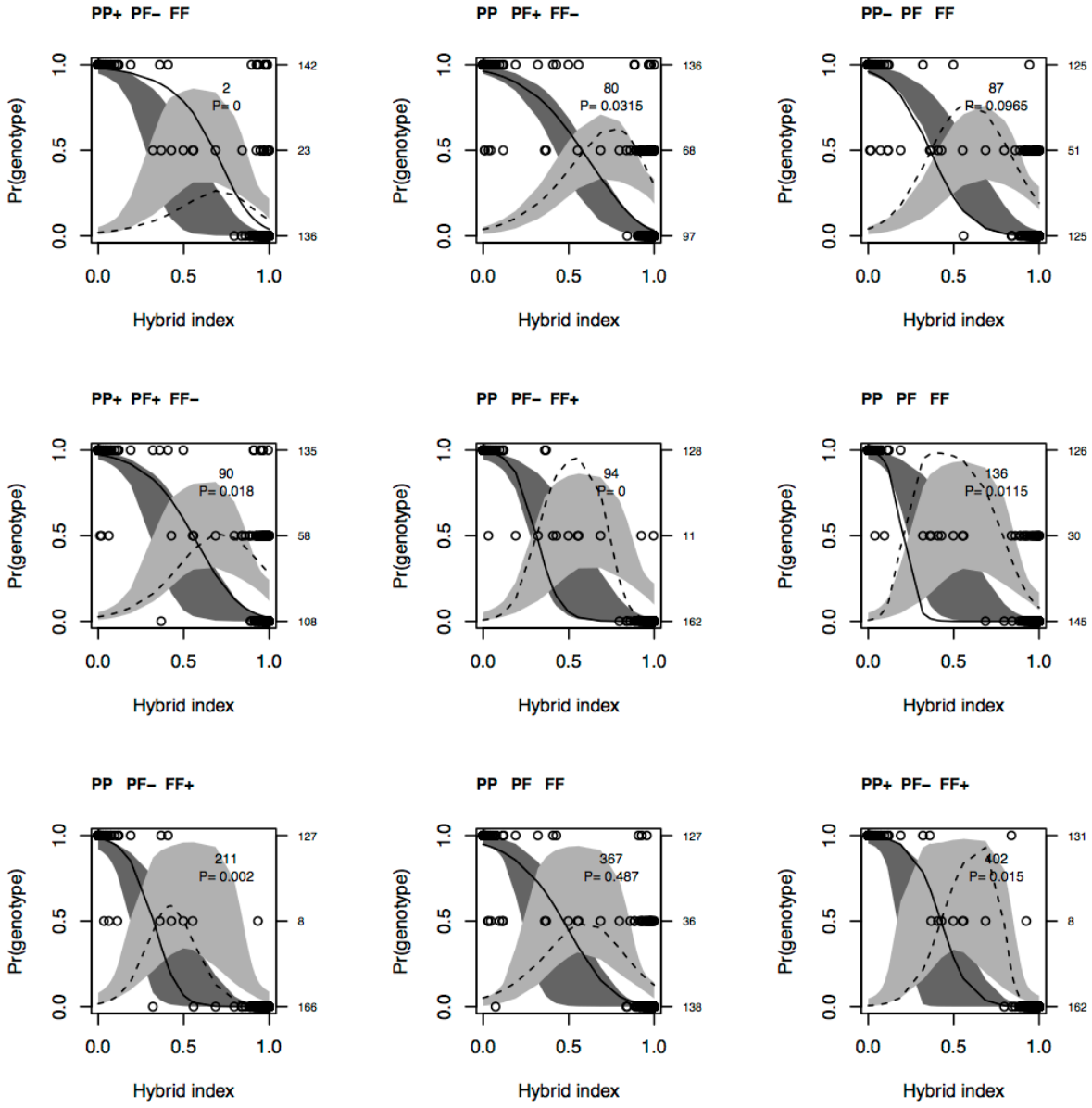

PP+ PF− FF
PP PF+ FF−
PP− PF FF
PP+ PF+ FF−
PP PF− FF+
PP PF FF
PP PF− FF+
PP PF FF
PP+ PF− FF+
Pr(genotype)
Hybrid index
2
P= 0
80
P= 0.0315
87
P= 0.0965
90
P= 0.018
94
P= 0
136
P= 0.0115
211
P= 0.002
367
P= 0.487
402
P= 0.015

Figure S5.1A (Continued)

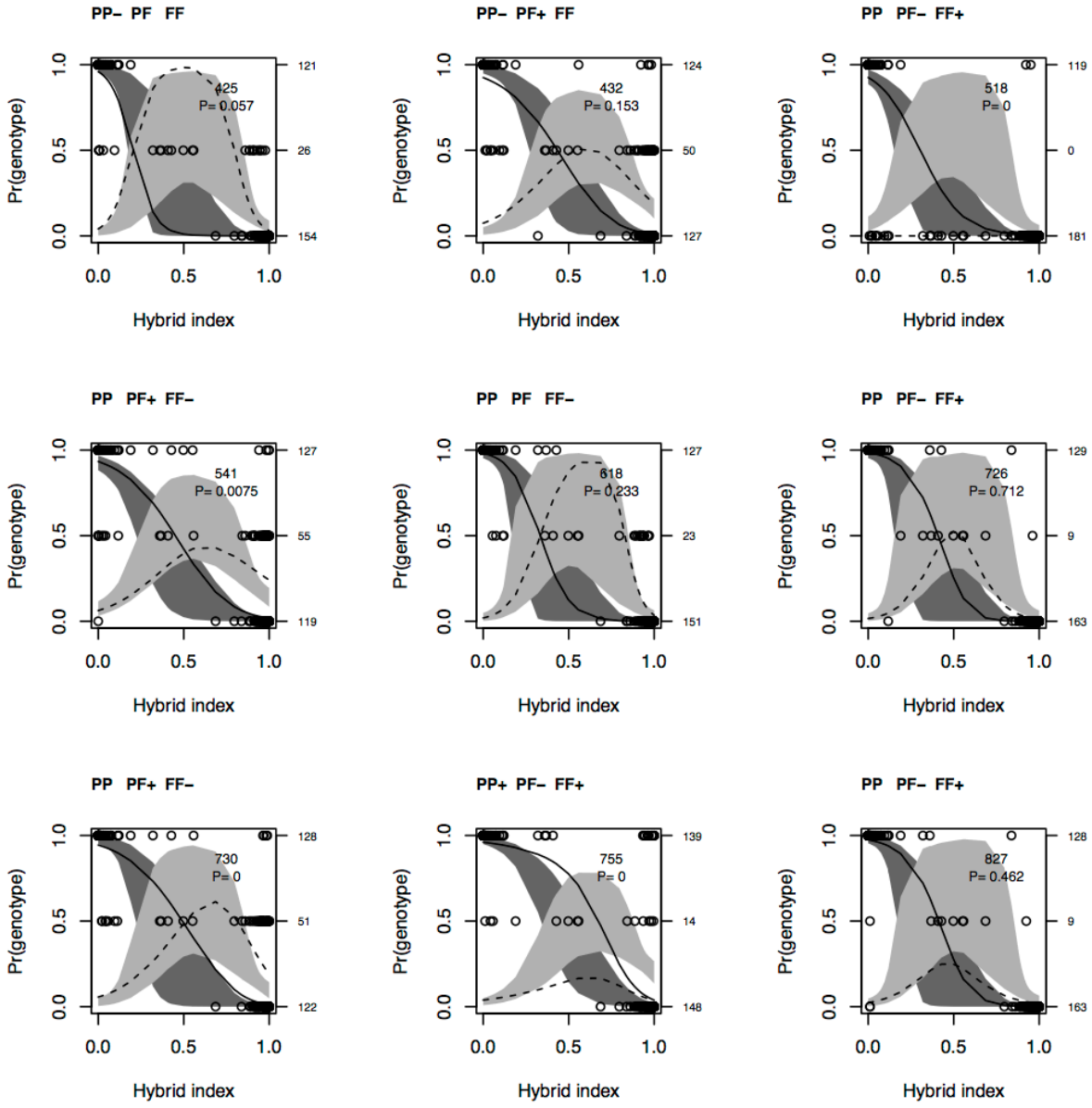


Figure S5.1A (Continued)

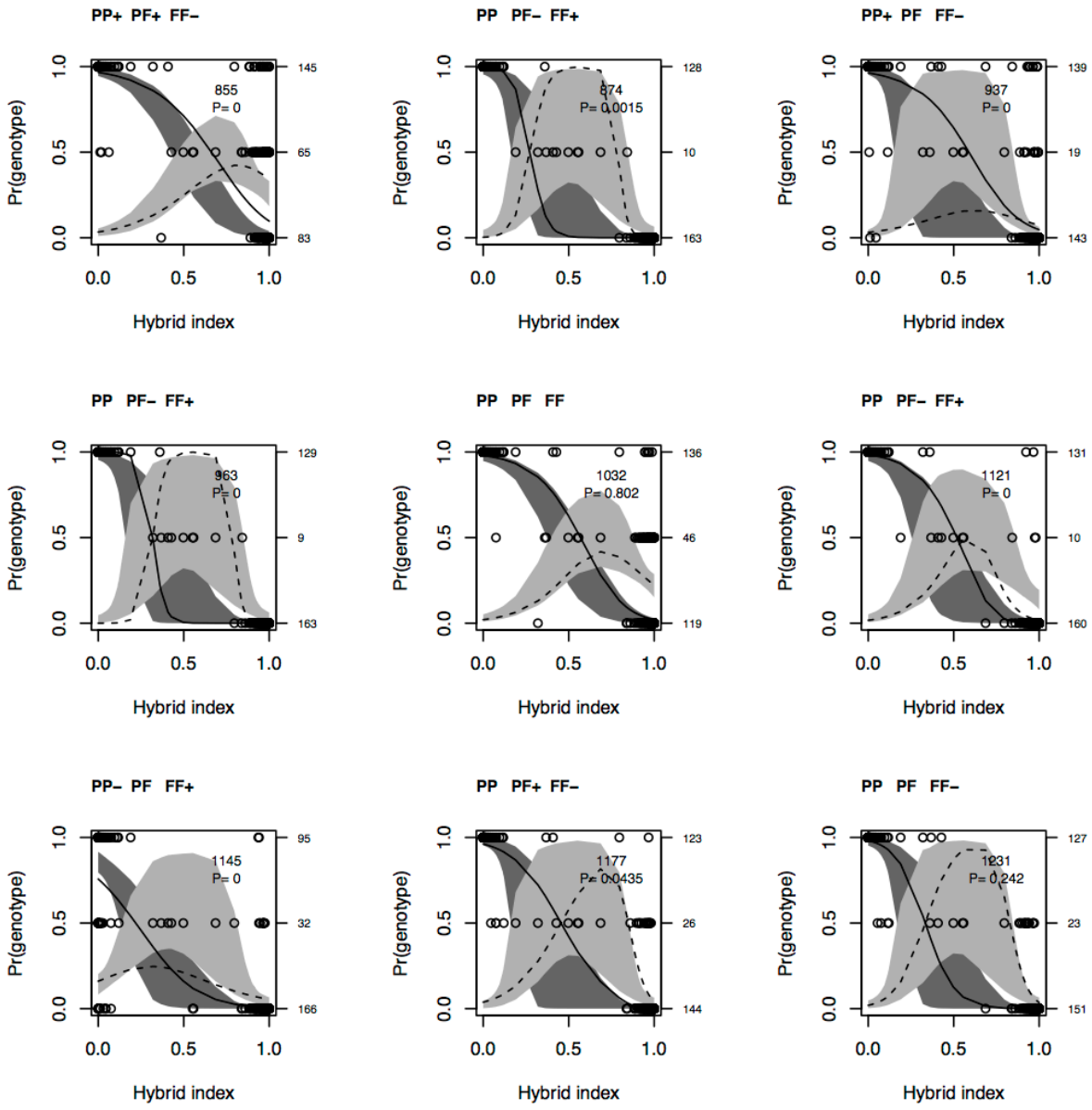


Figure S5.1A (Continued)

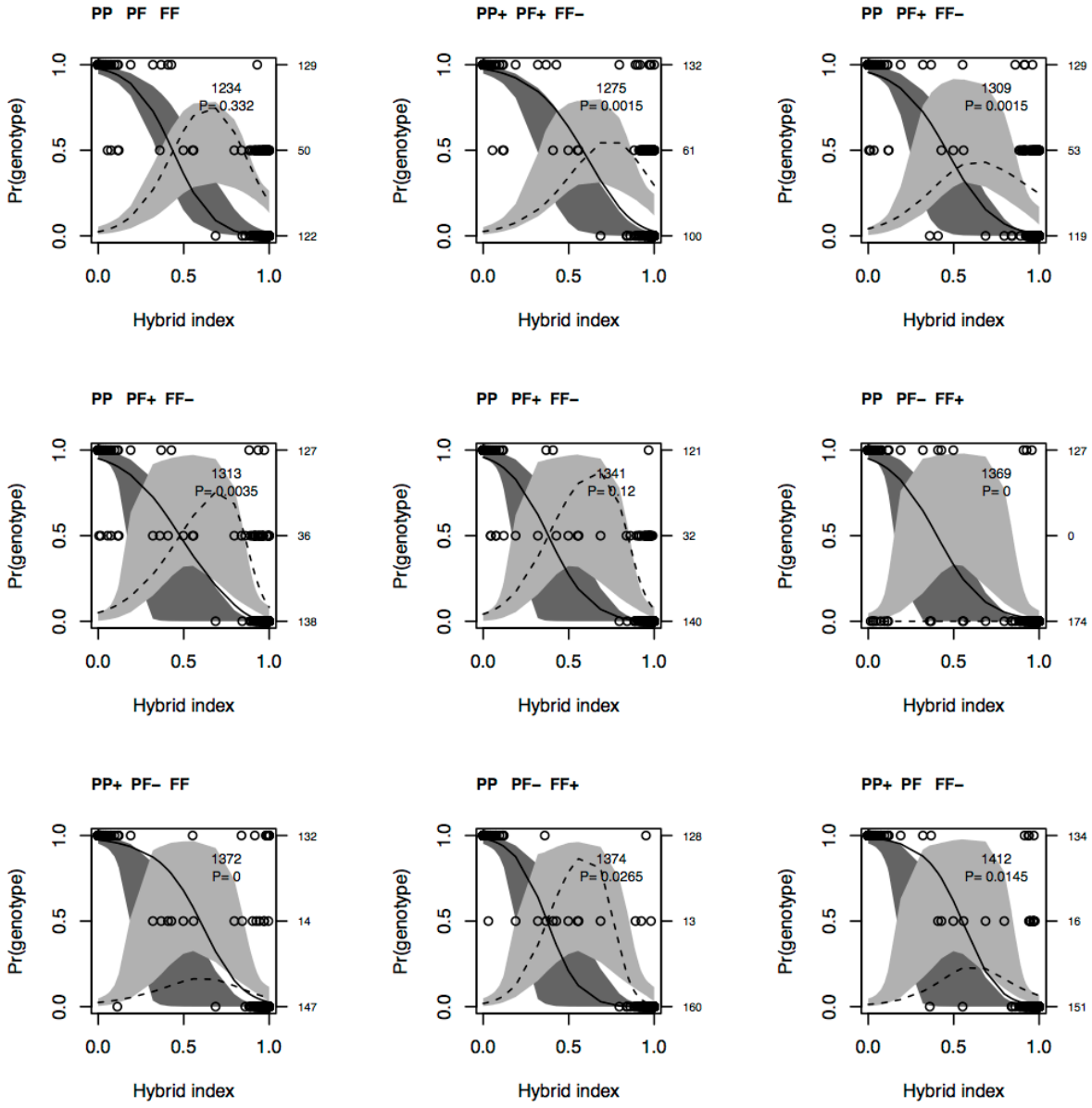


Figure S5.1A (Continued)

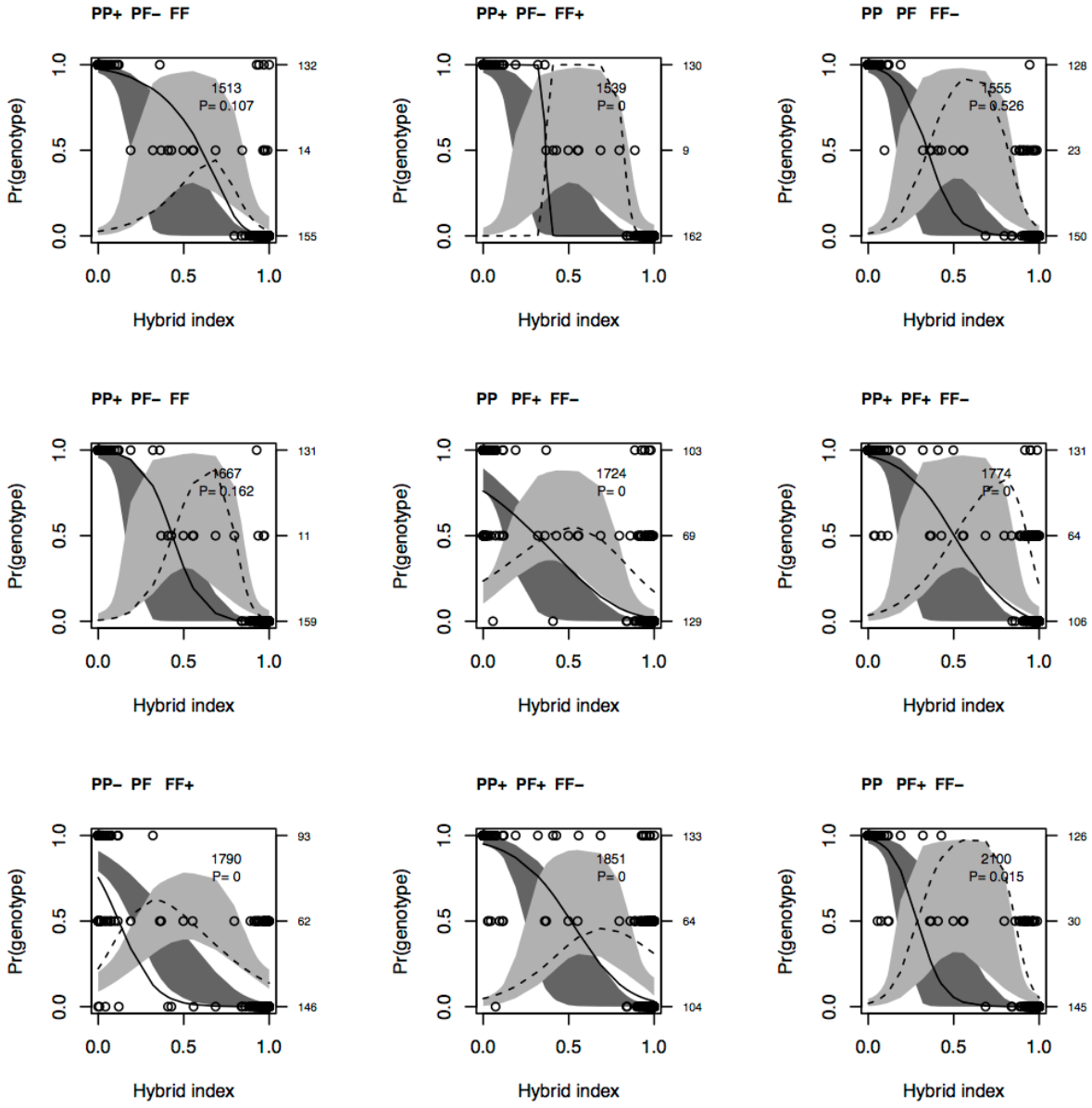


Figure S5.1A (Continued)

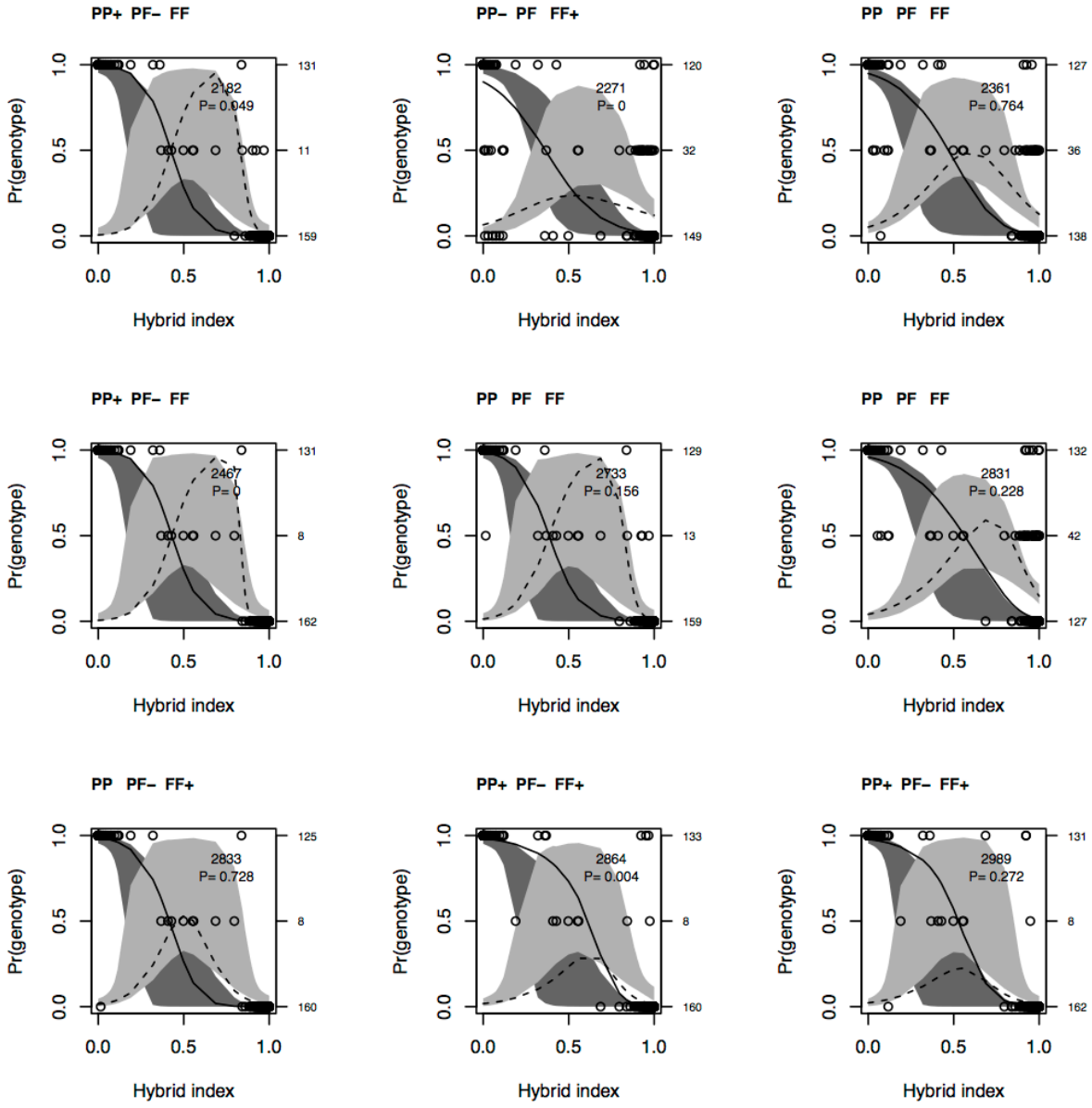


Figure S5.1A (Continued)

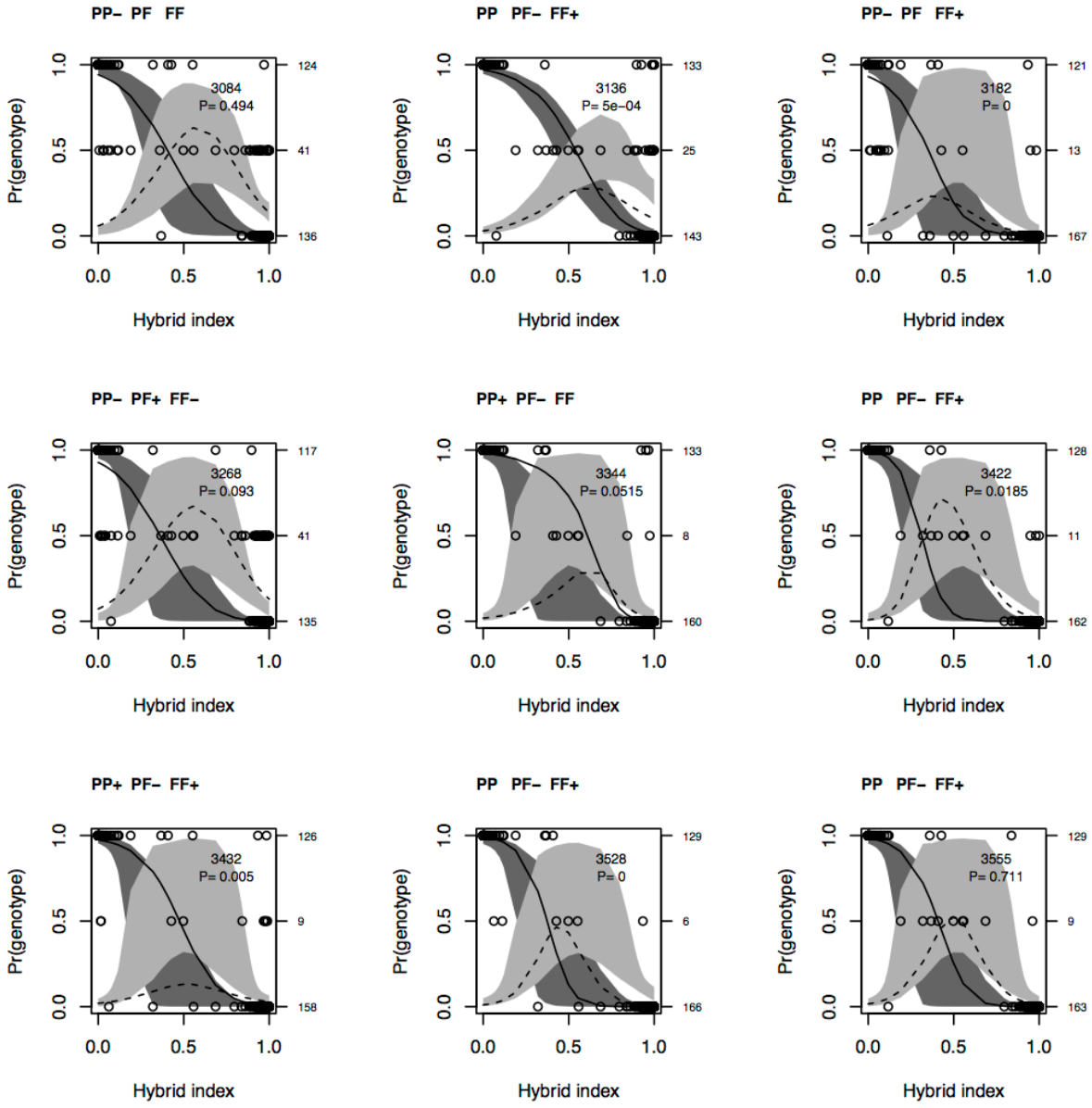


Figure S5.1A (Continued)

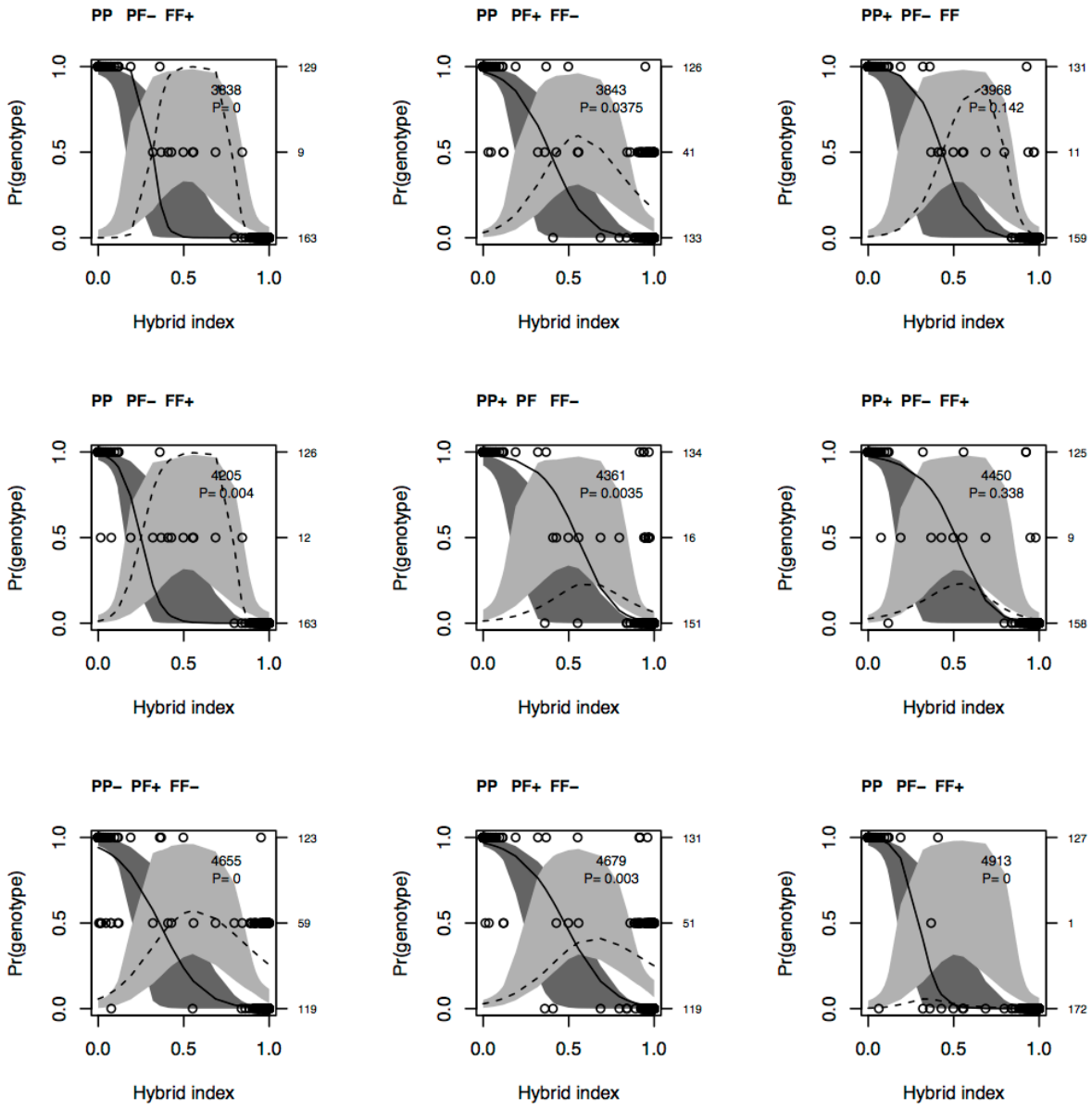


Figure S5.1A (Continued)

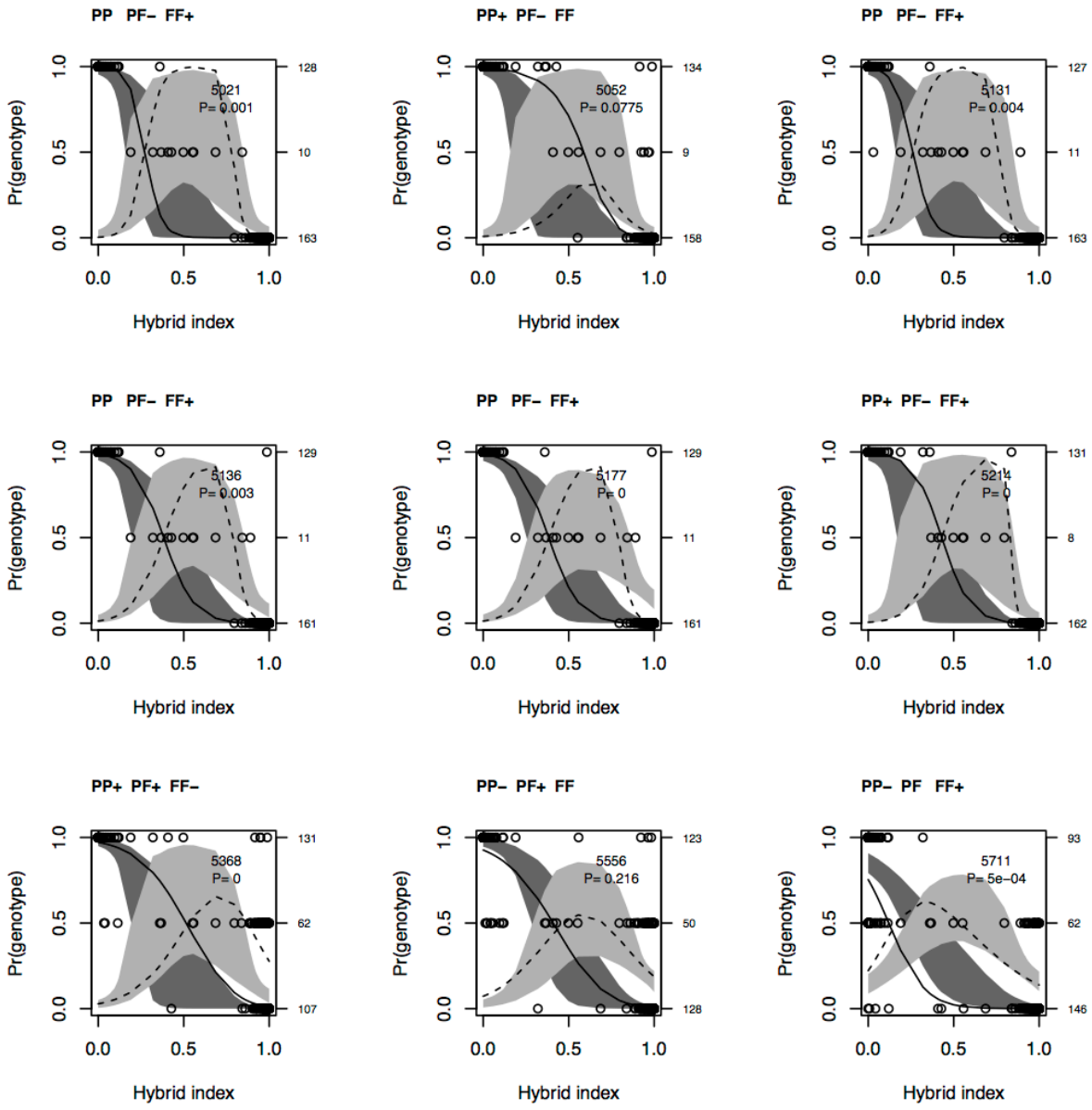


Figure S5.1A (Continued)

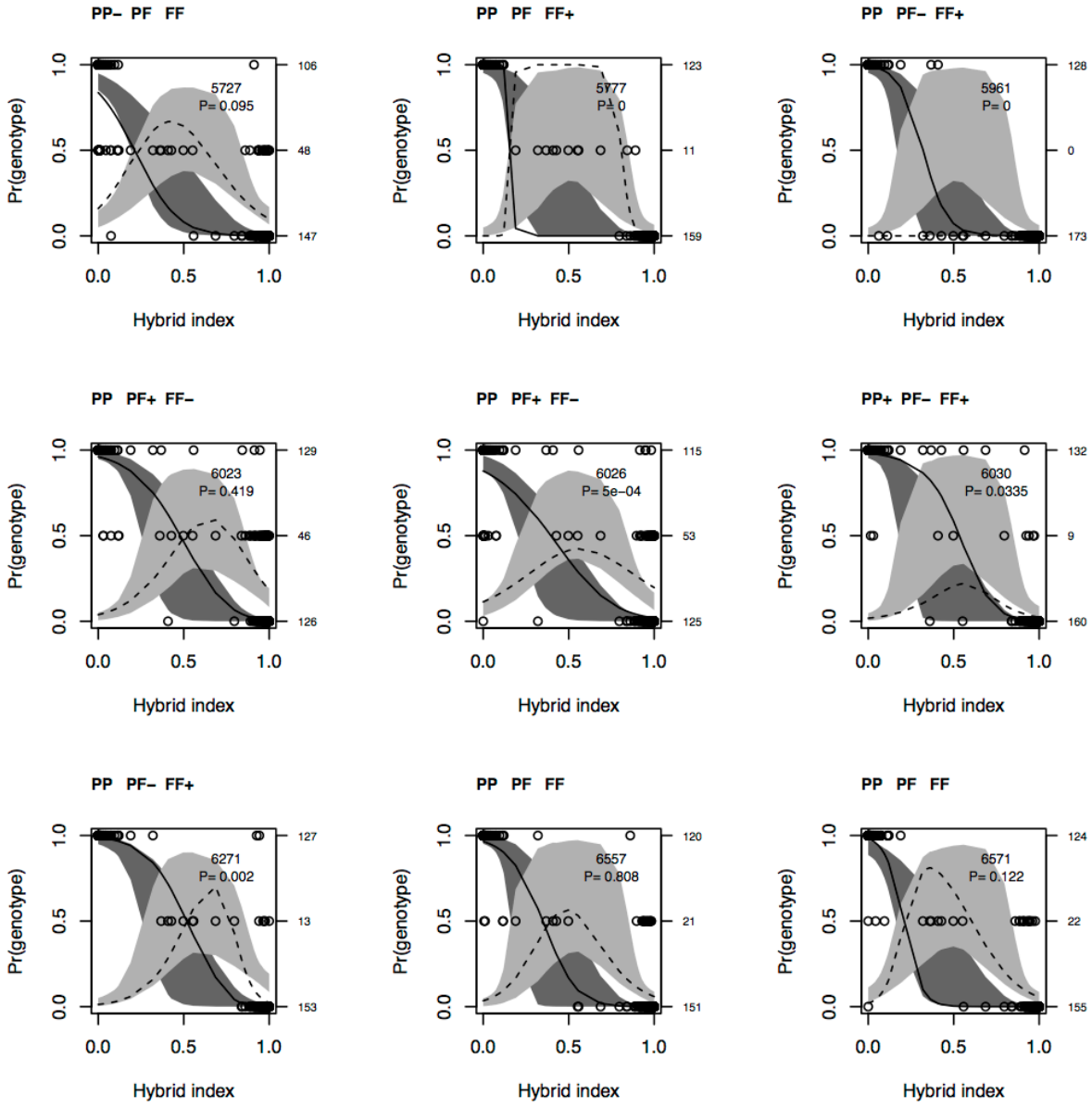


Figure S5.1A (Continued)

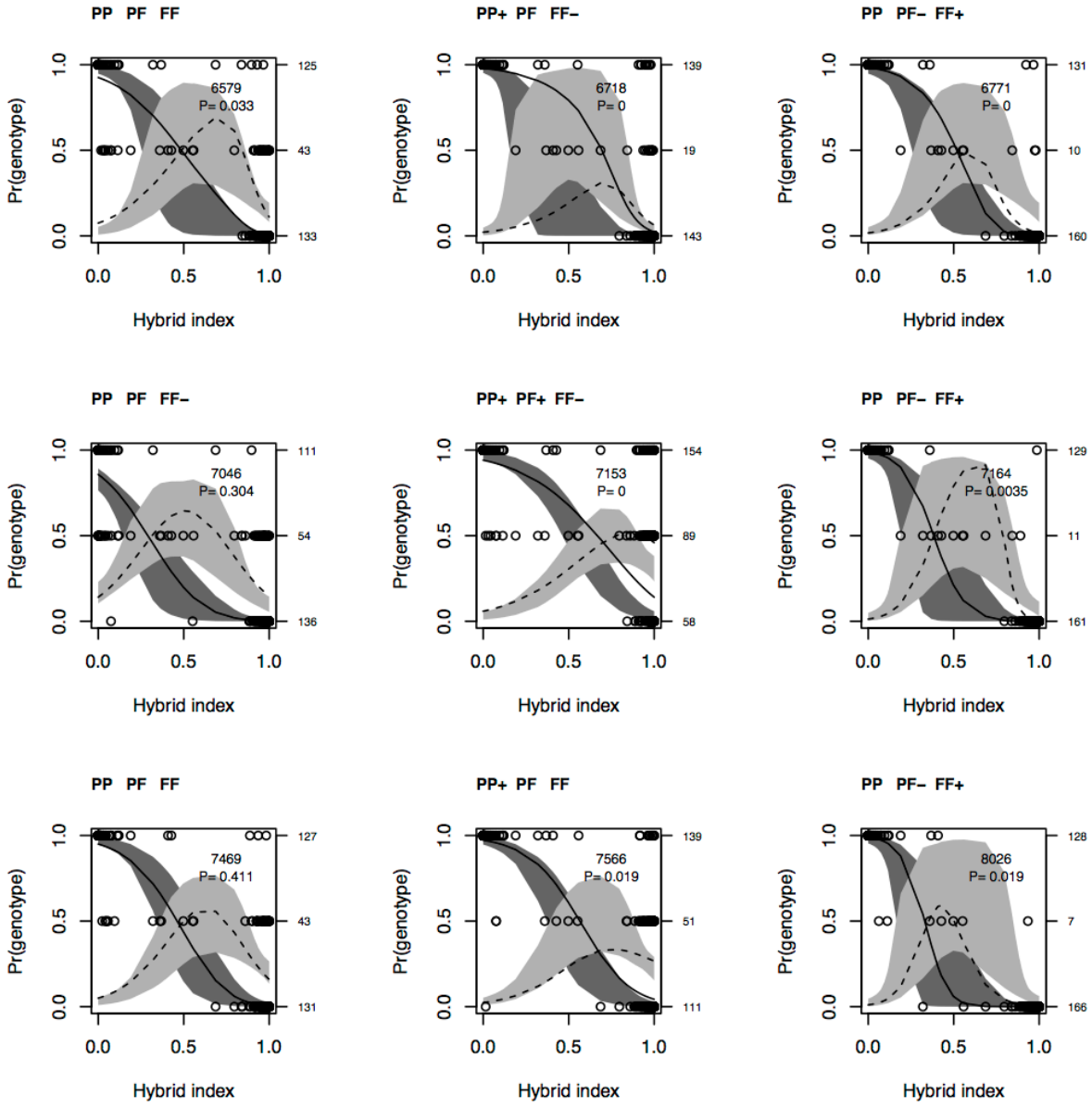


Figure S5.1A (Continued)

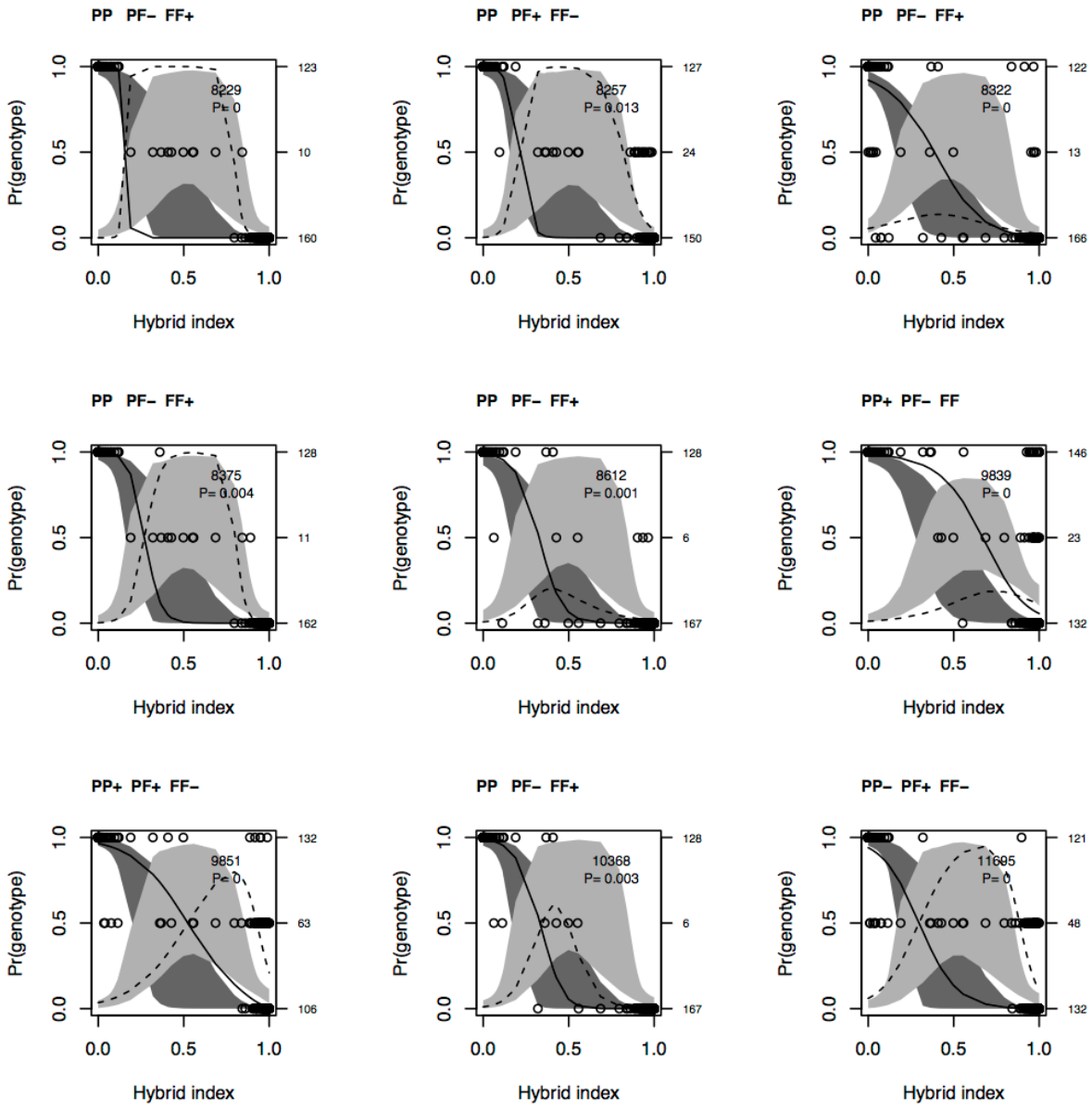


Figure S5.1A (Continued)

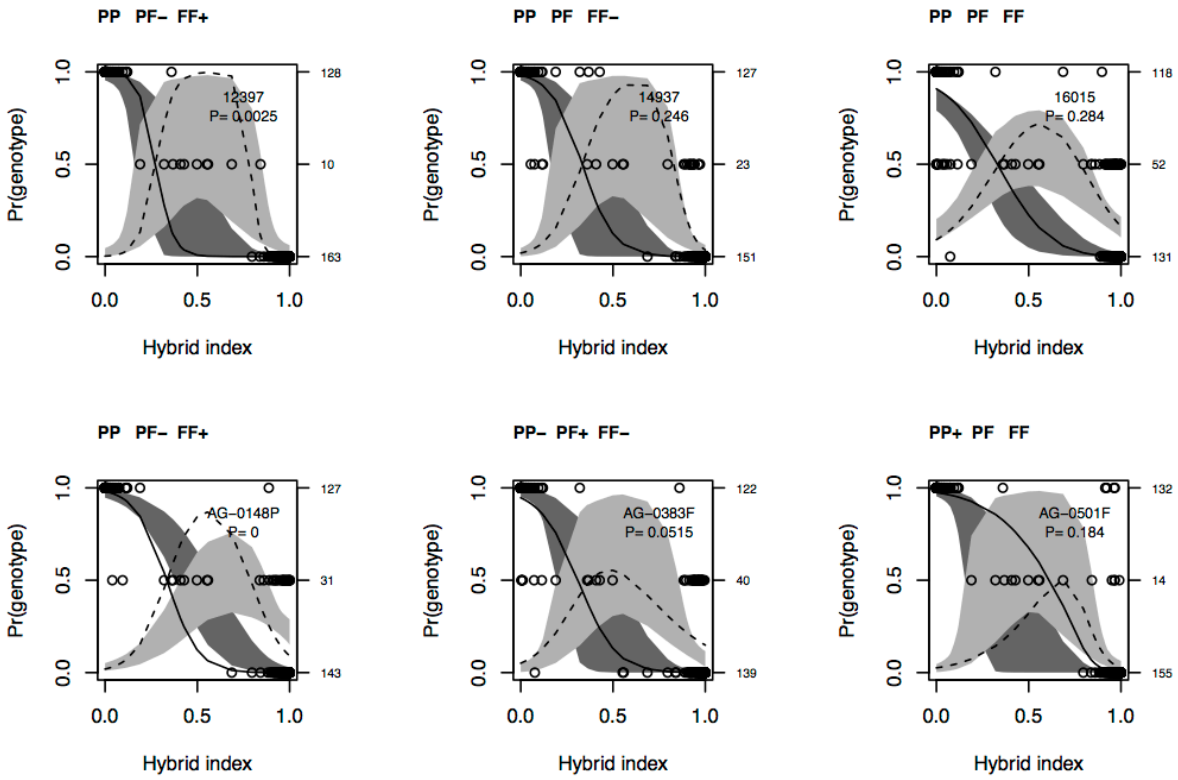


Figure S5.1B (Continued)

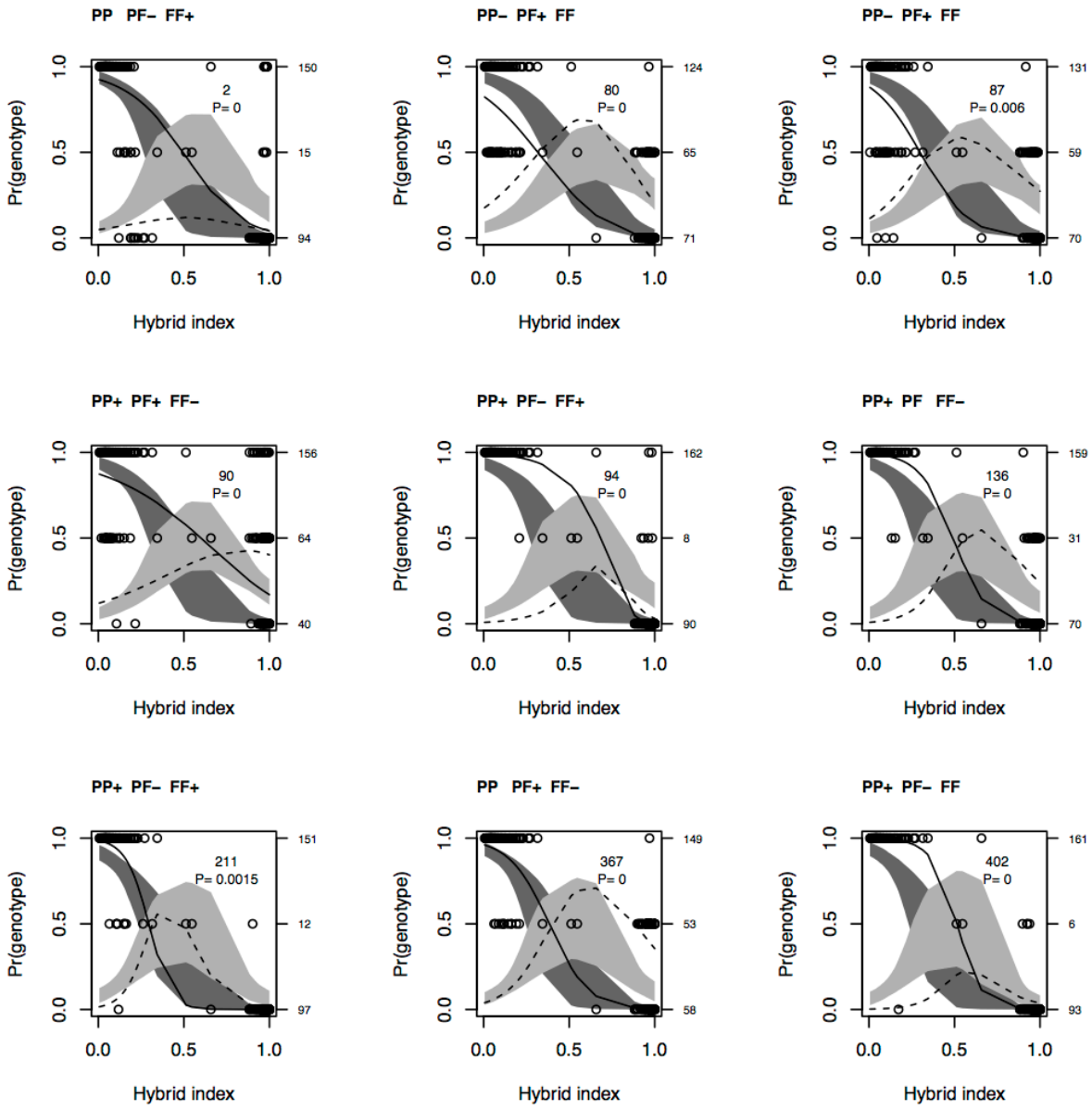


Figure S5.1B (Continued)

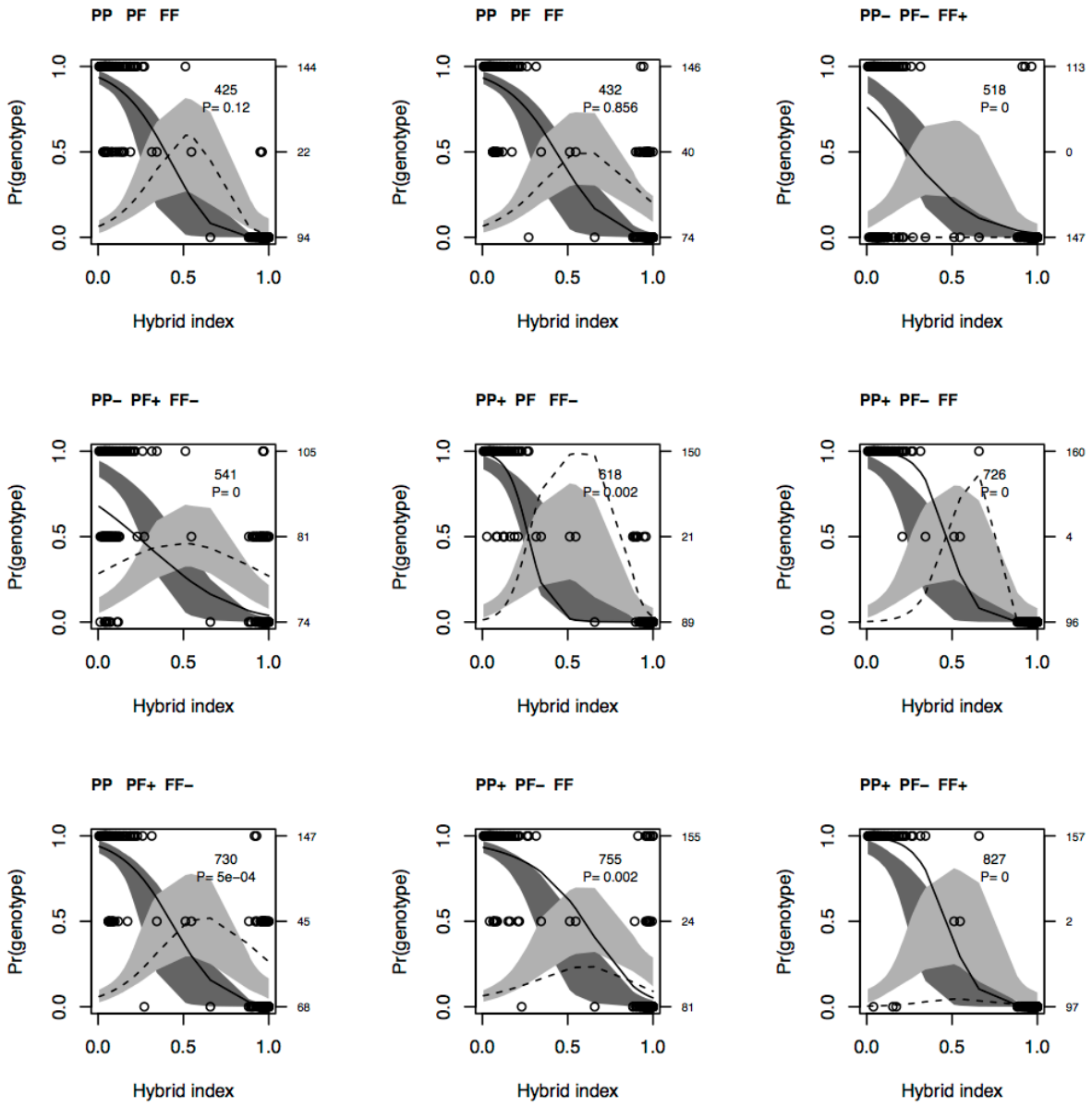


Figure S5.1B (Continued)

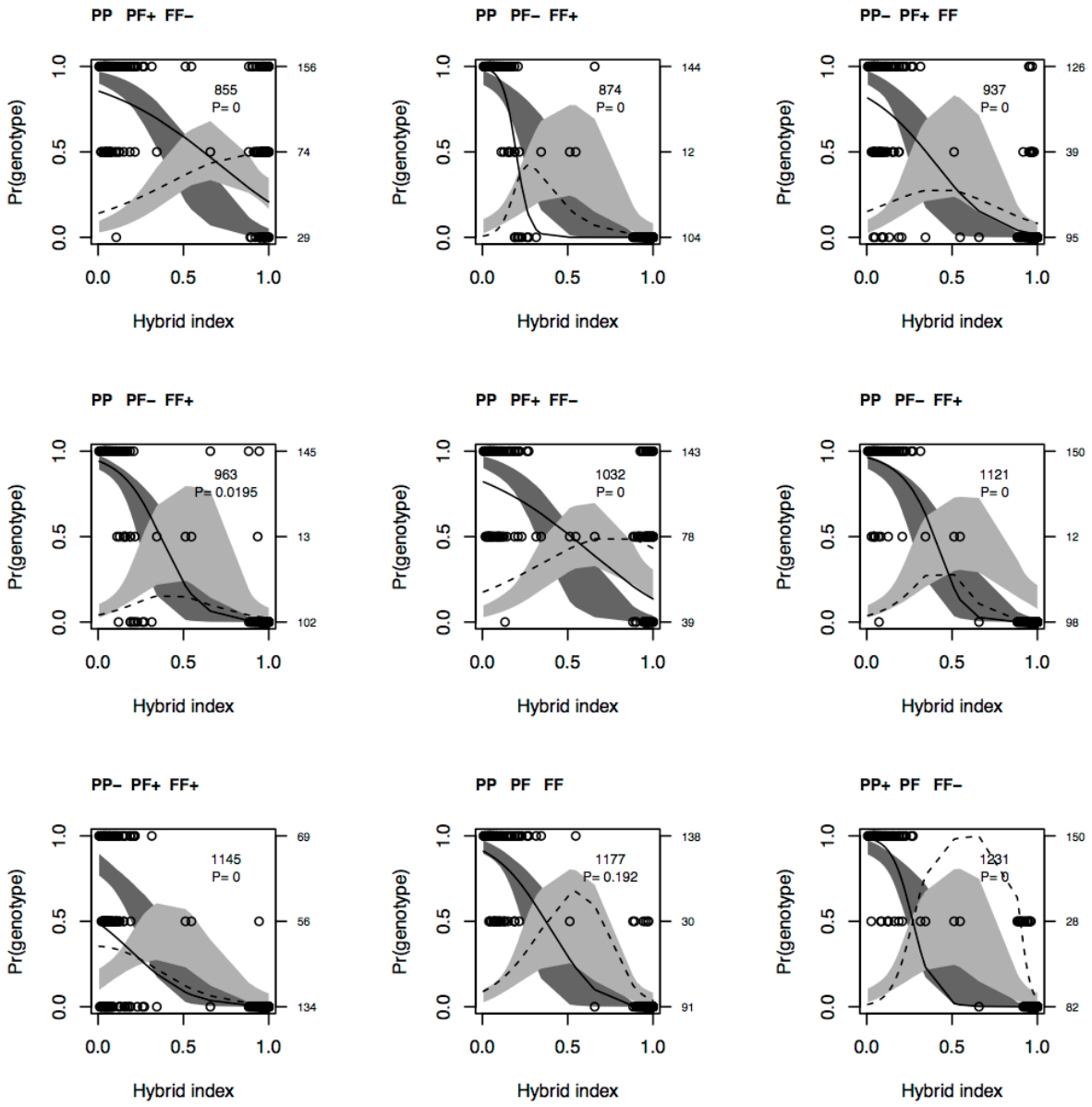


Figure S5.1B (Continued)

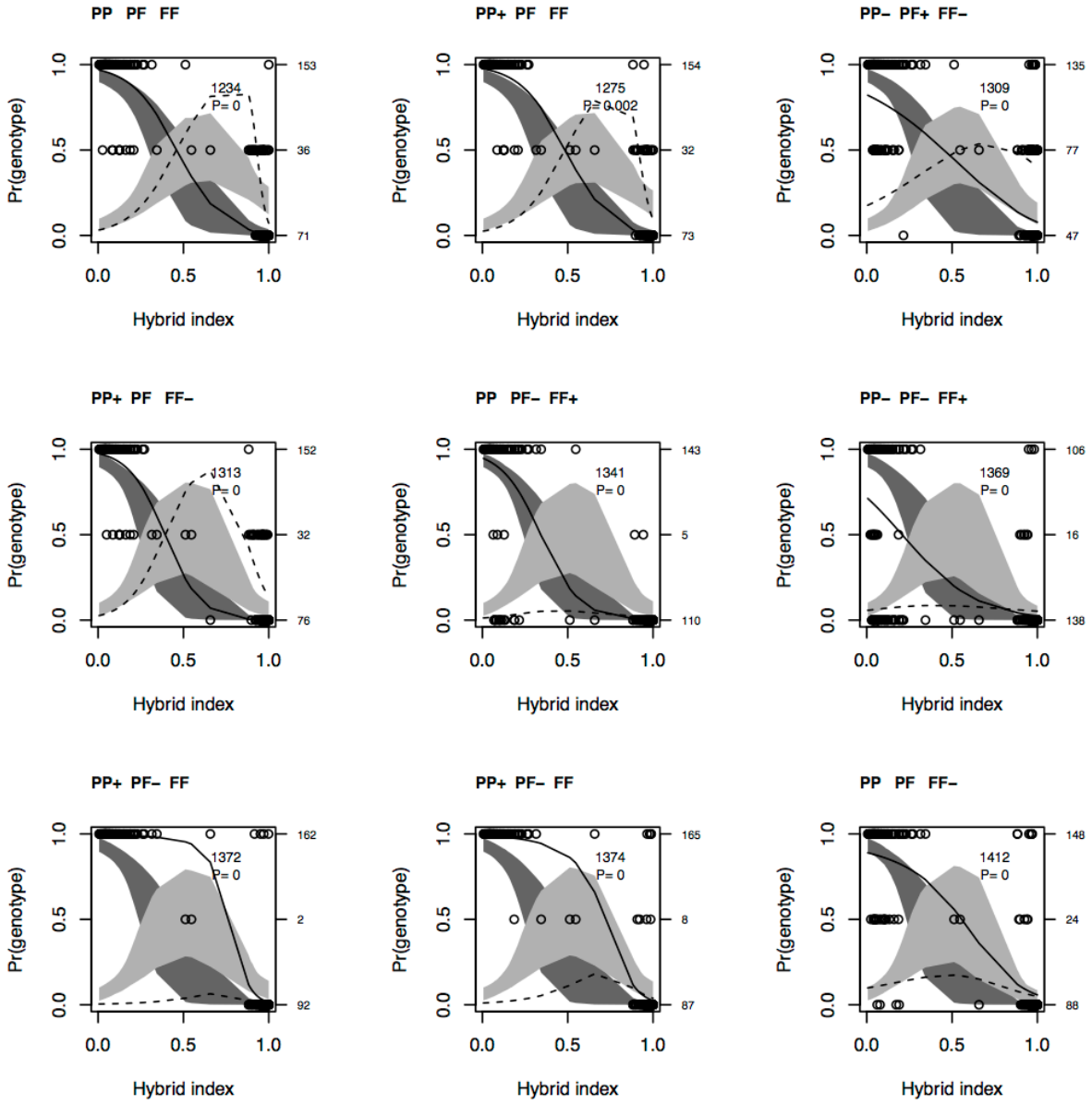


Figure S5.1B (Continued)

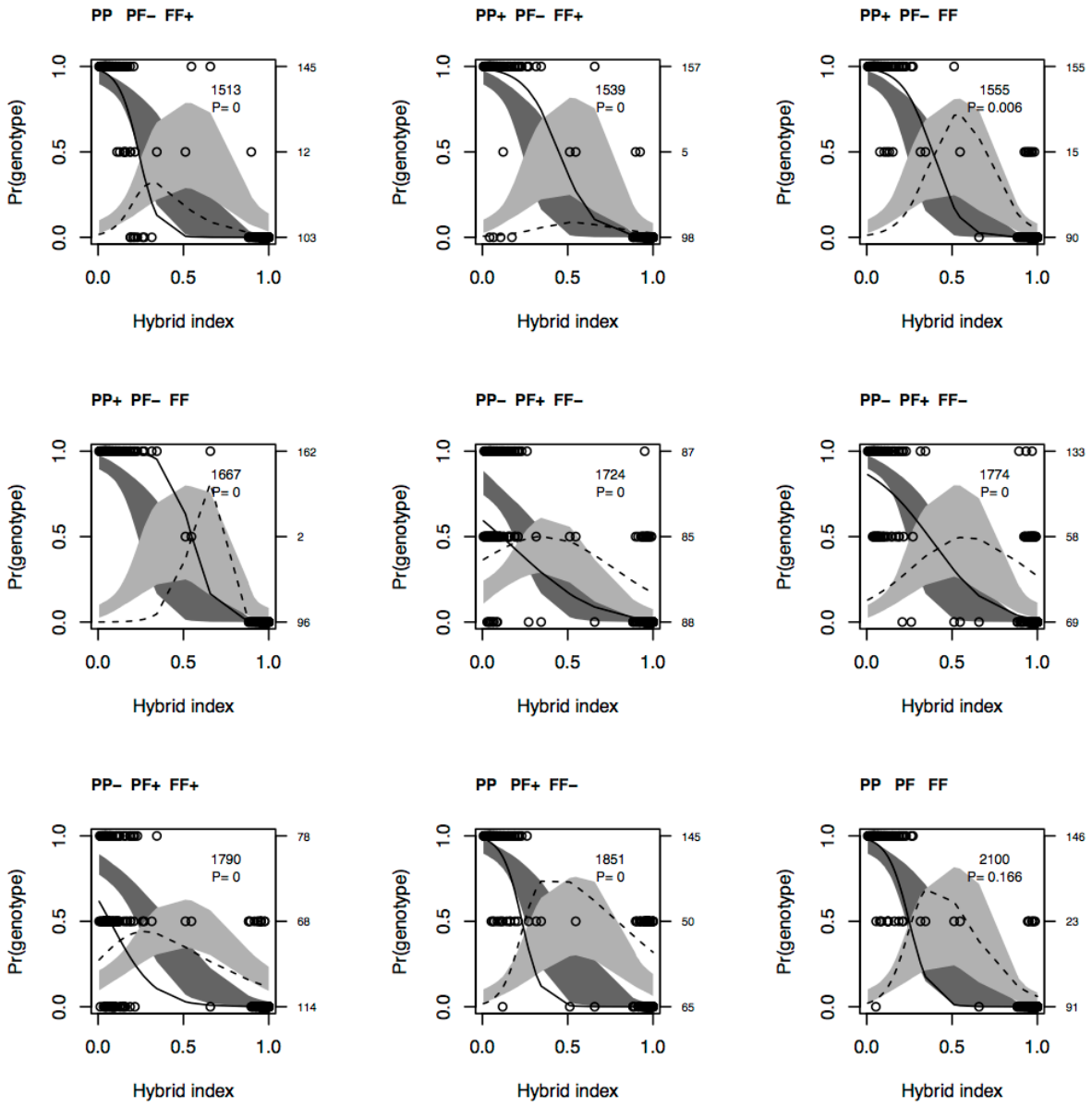


Figure S5.1B (Continued)

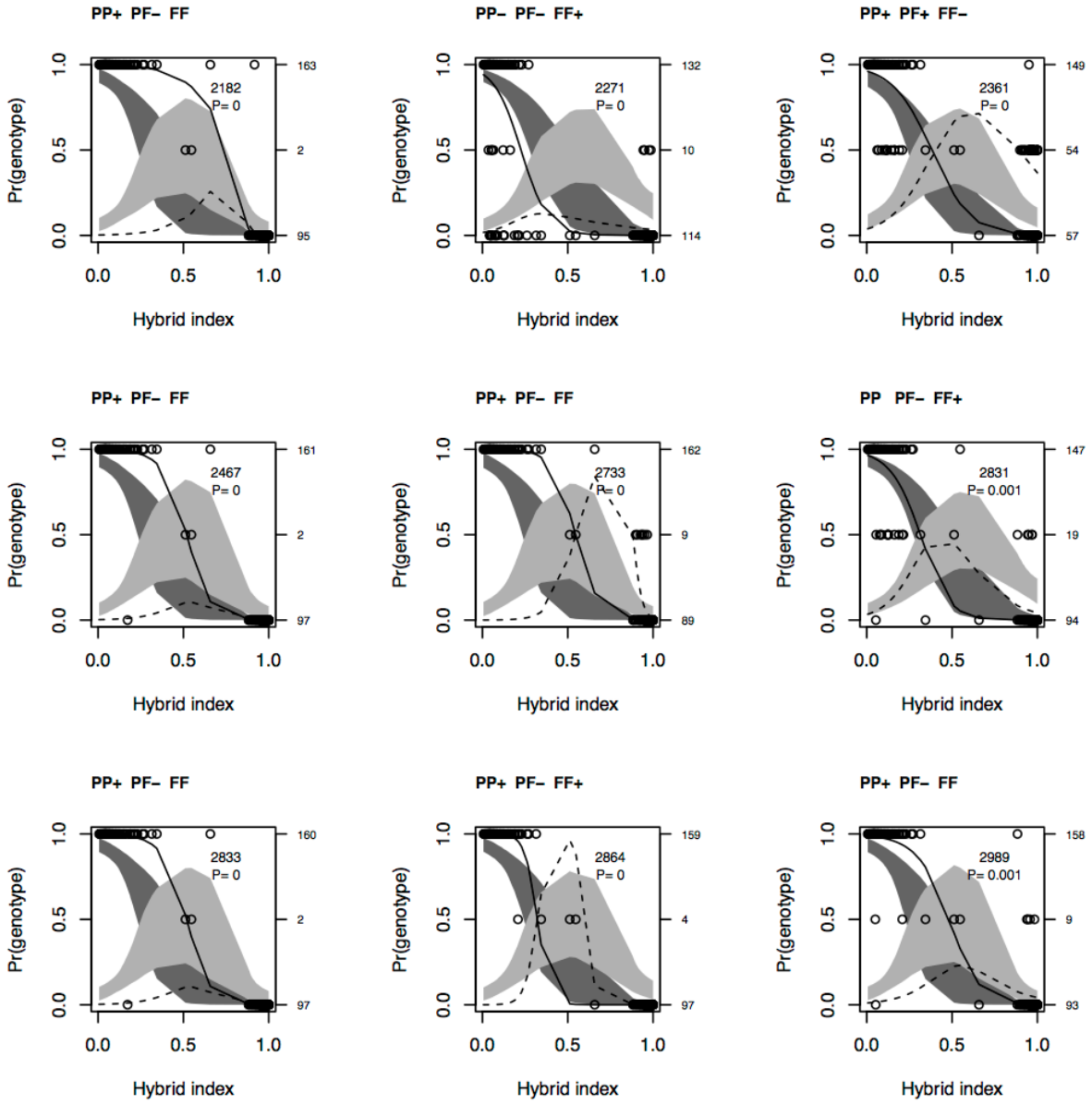


Figure S5.1B (Continued)

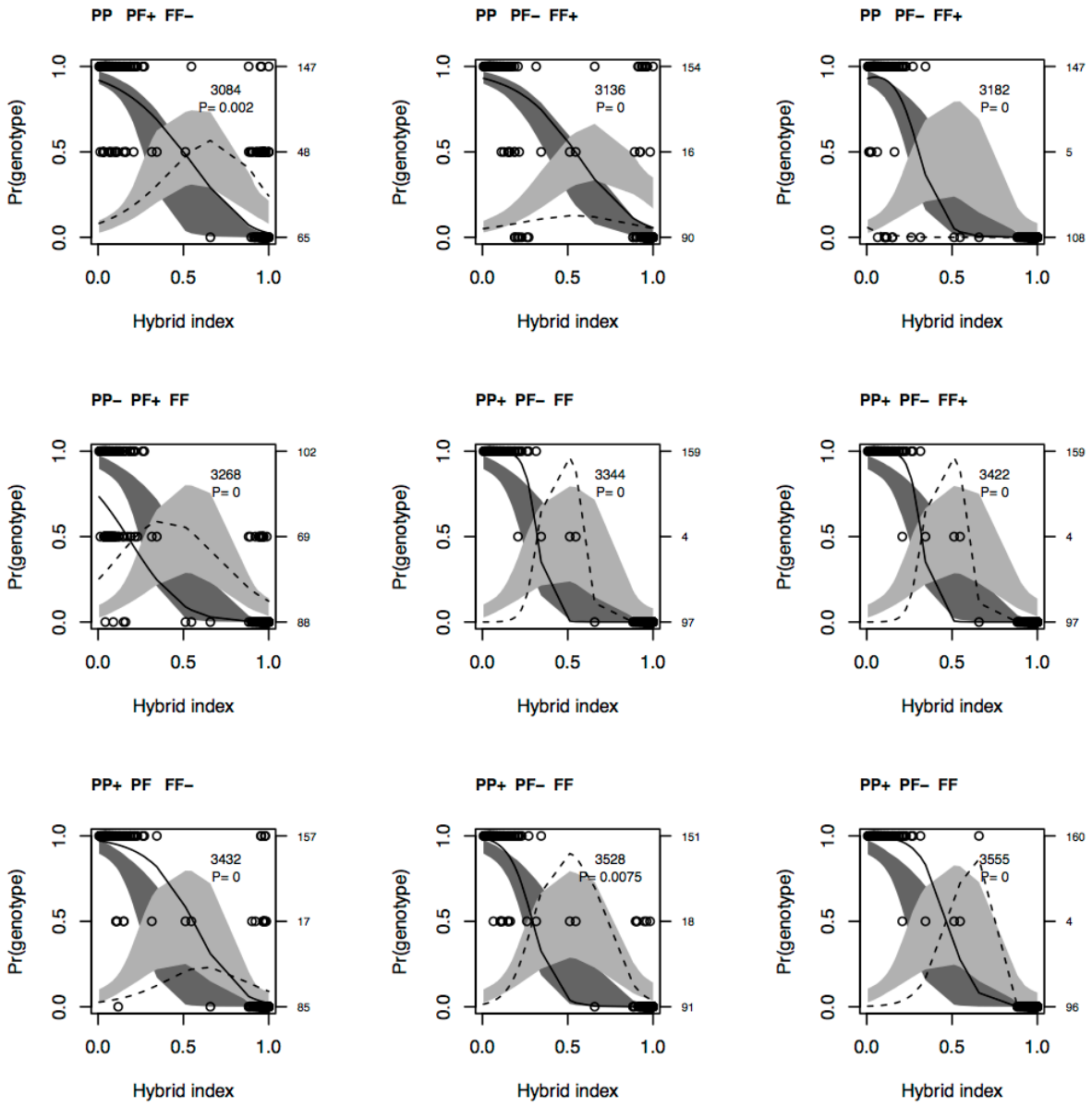


Figure S5.1B (Continued)

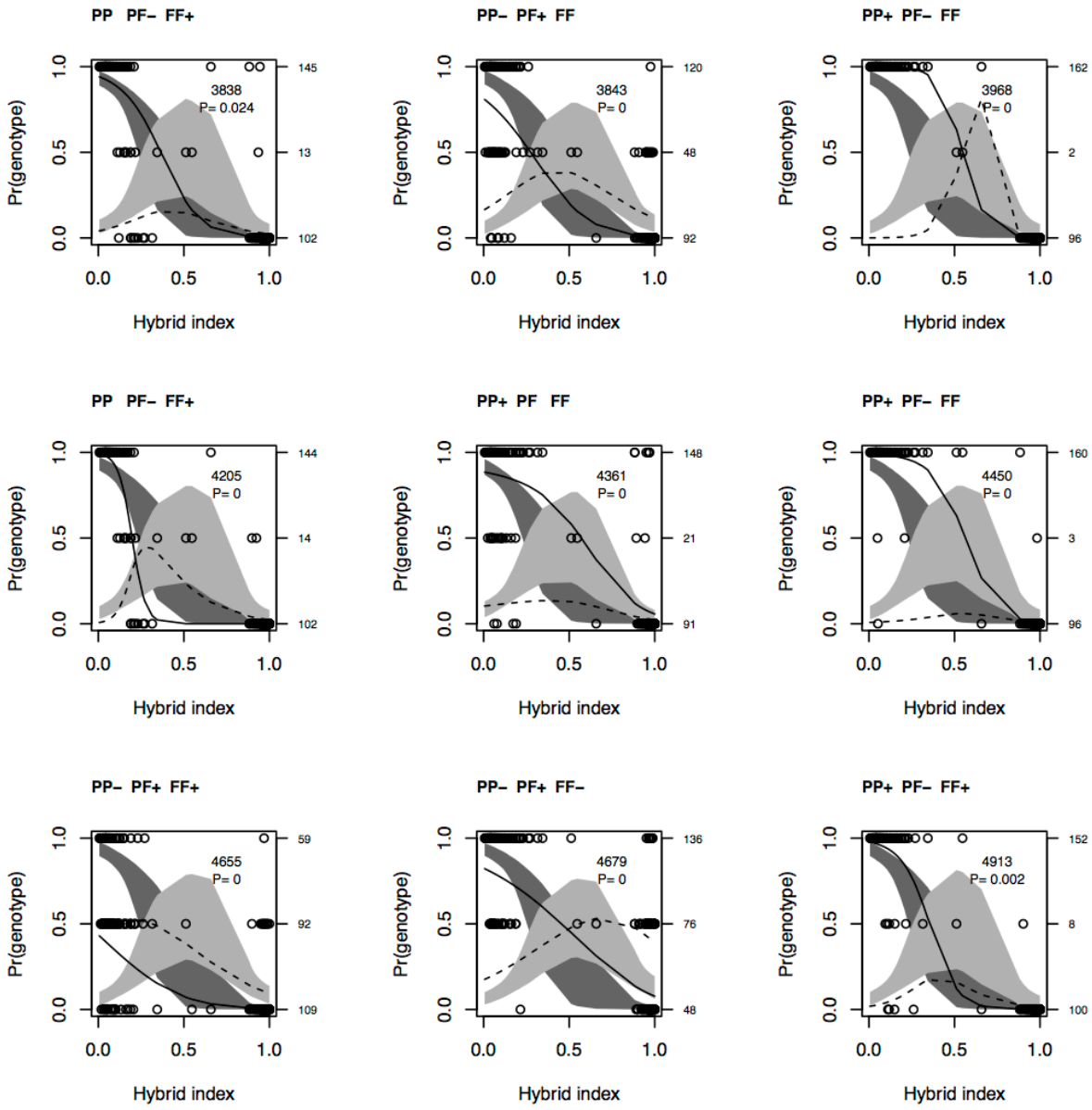


Figure S5.1B (Continued)

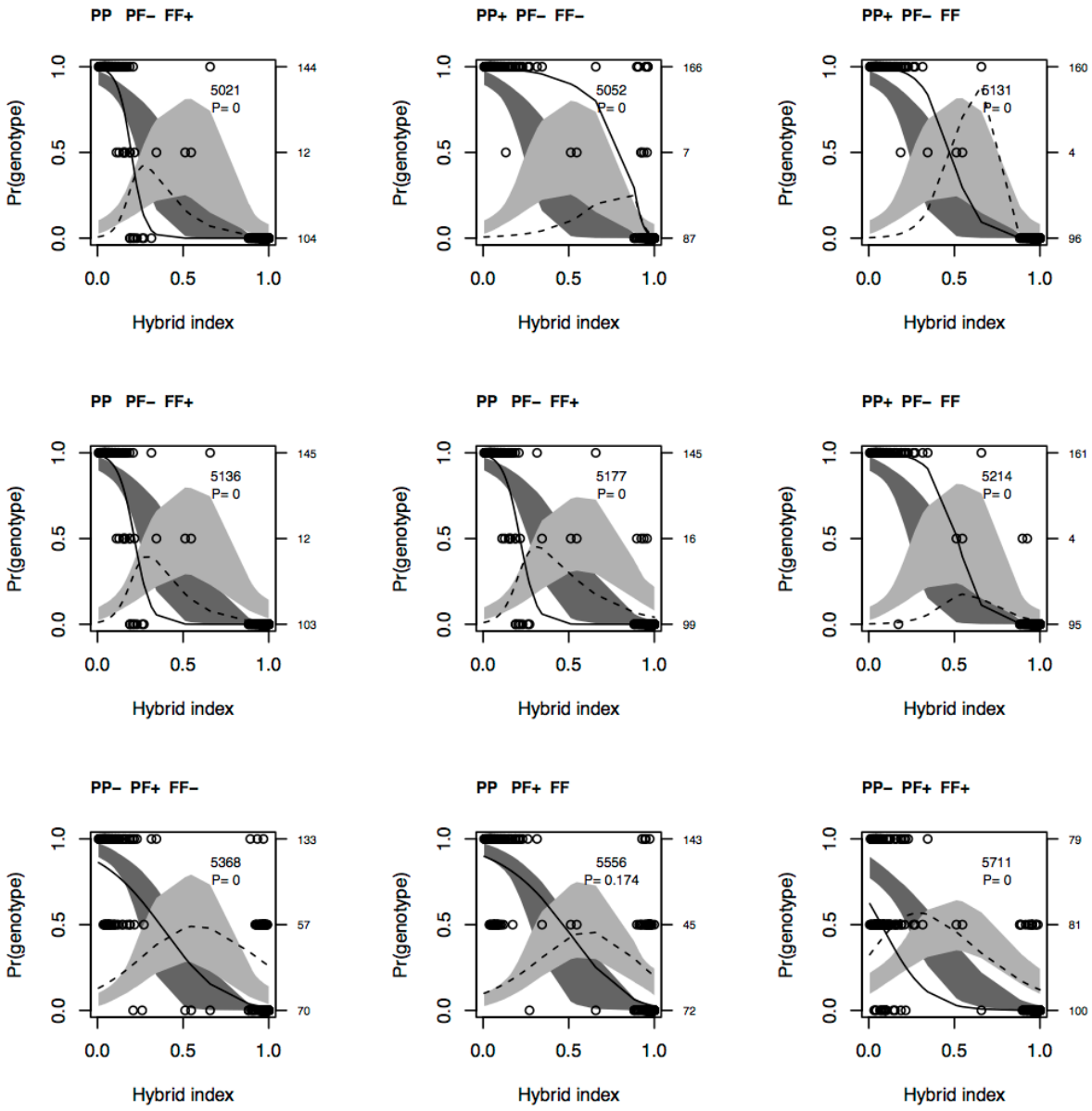


Figure S5.1B (Continued)

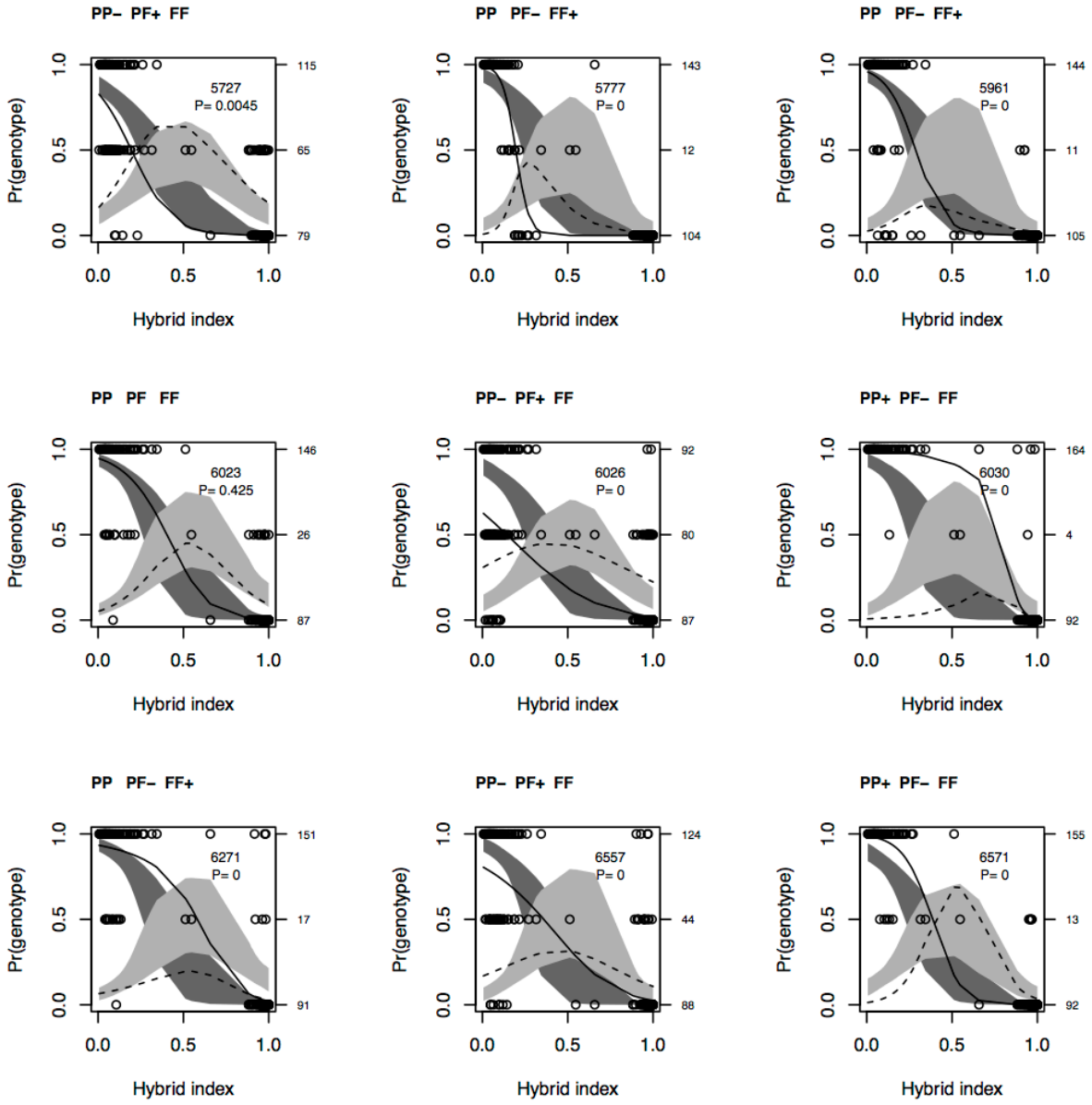


Figure S5.1B (Continued)

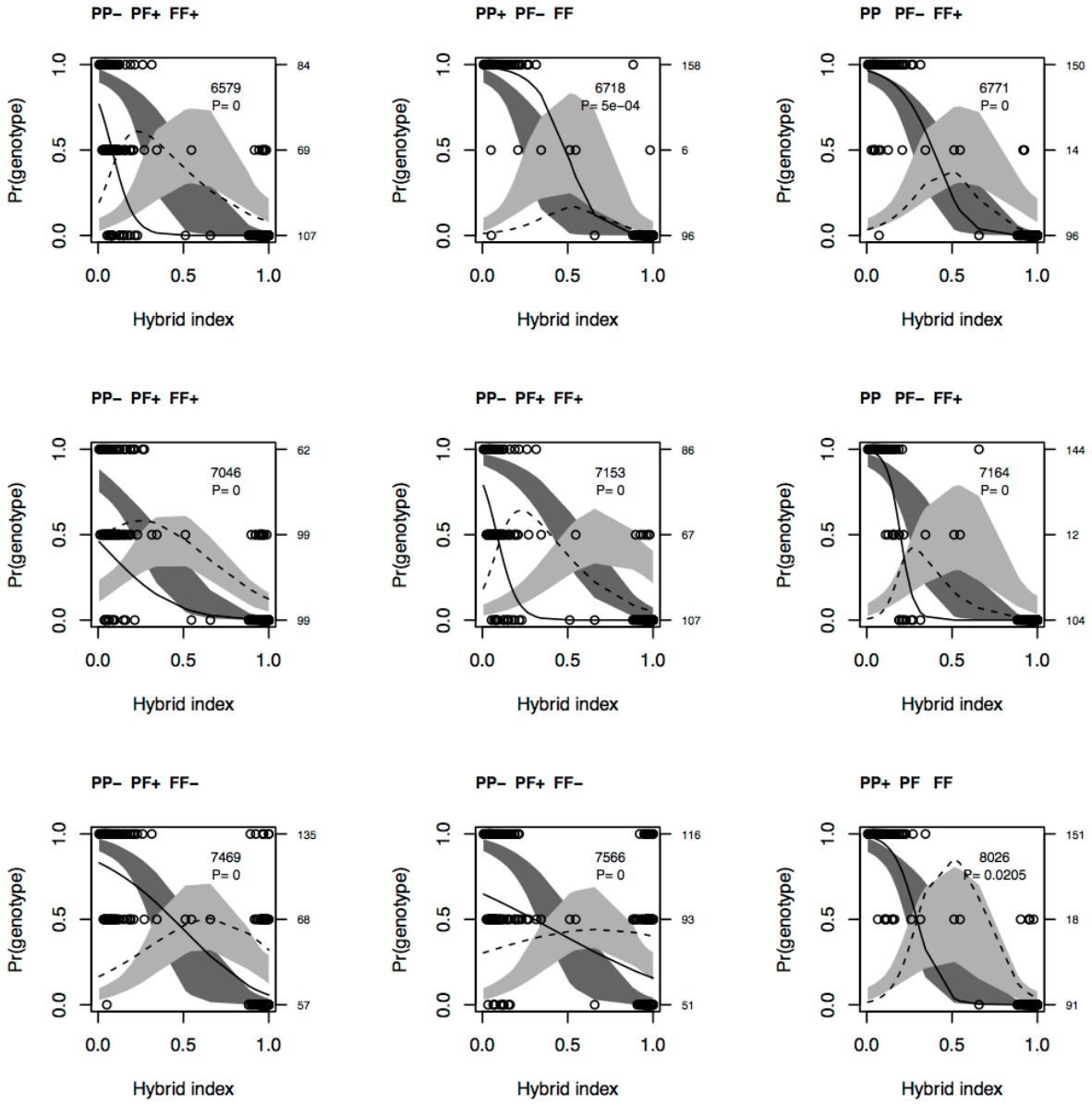


Figure S5.1B (Continued)

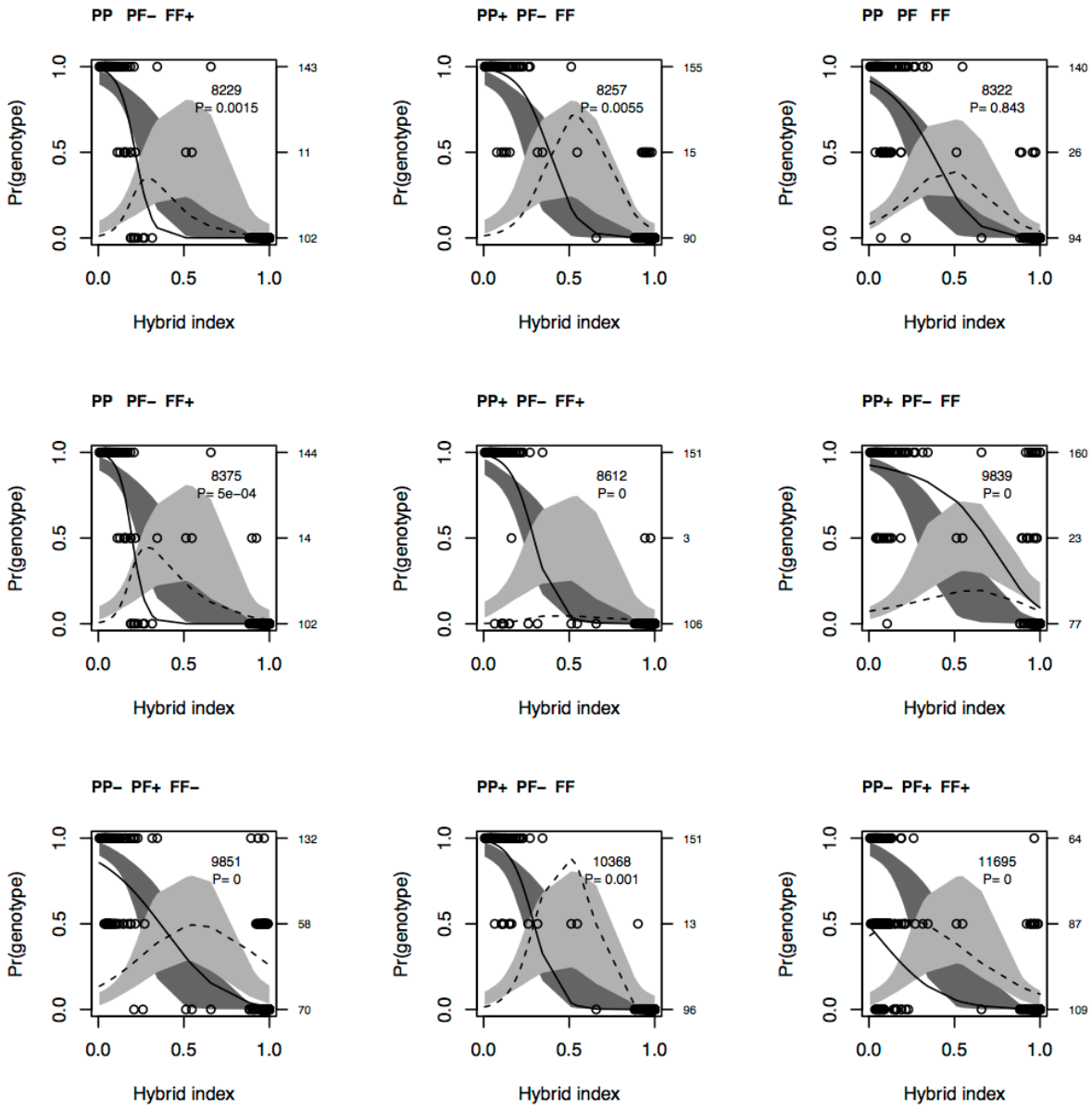
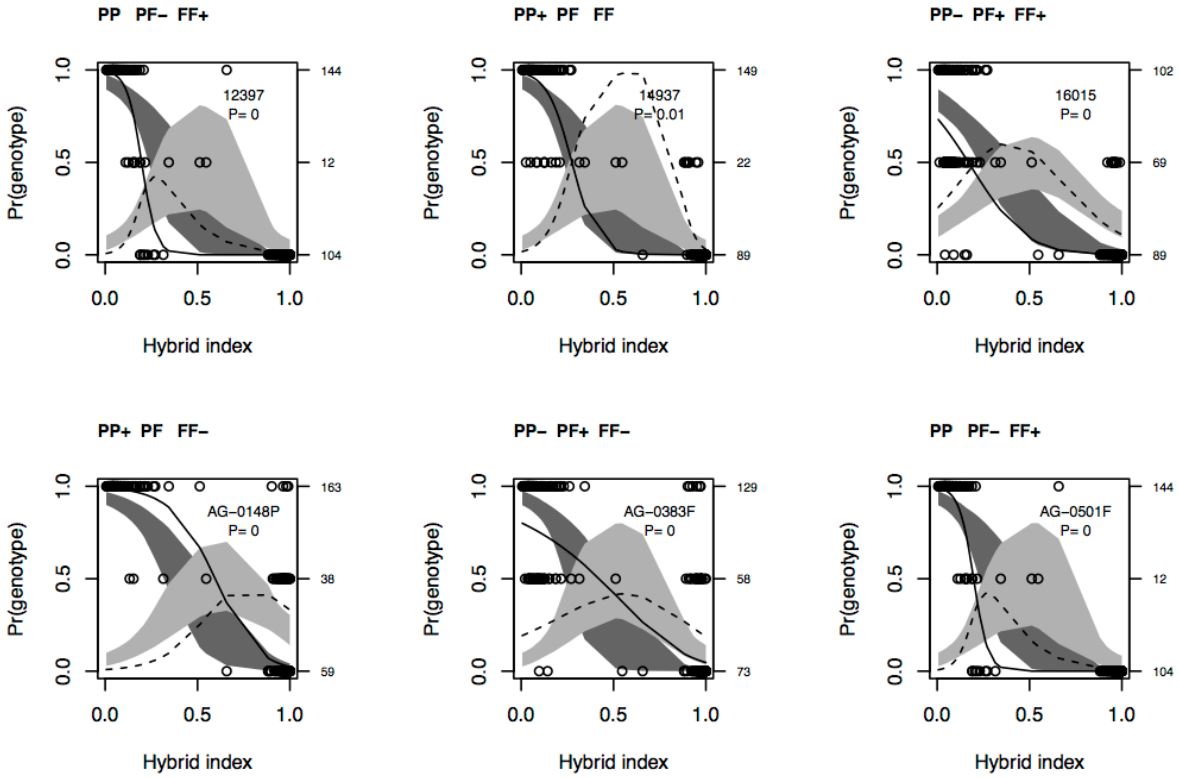


Figure S5.1B (Continued)



Figures S5.2. Individual geographic clines for mtDNA and three anonymous nuclear markers from Ross and Harrison (2002) and 114 SNP markers genotyped in Connecticut hybrid zone. Each geographic cline depicts the change in the frequency of the *G. pennsylvanicus* allele across from *G. pennsylvanicus* habitat (loamy soil, 0 m) to *G. firmus* habitat (sandy soil, 500 m).

Figure S5.2 (Continued)

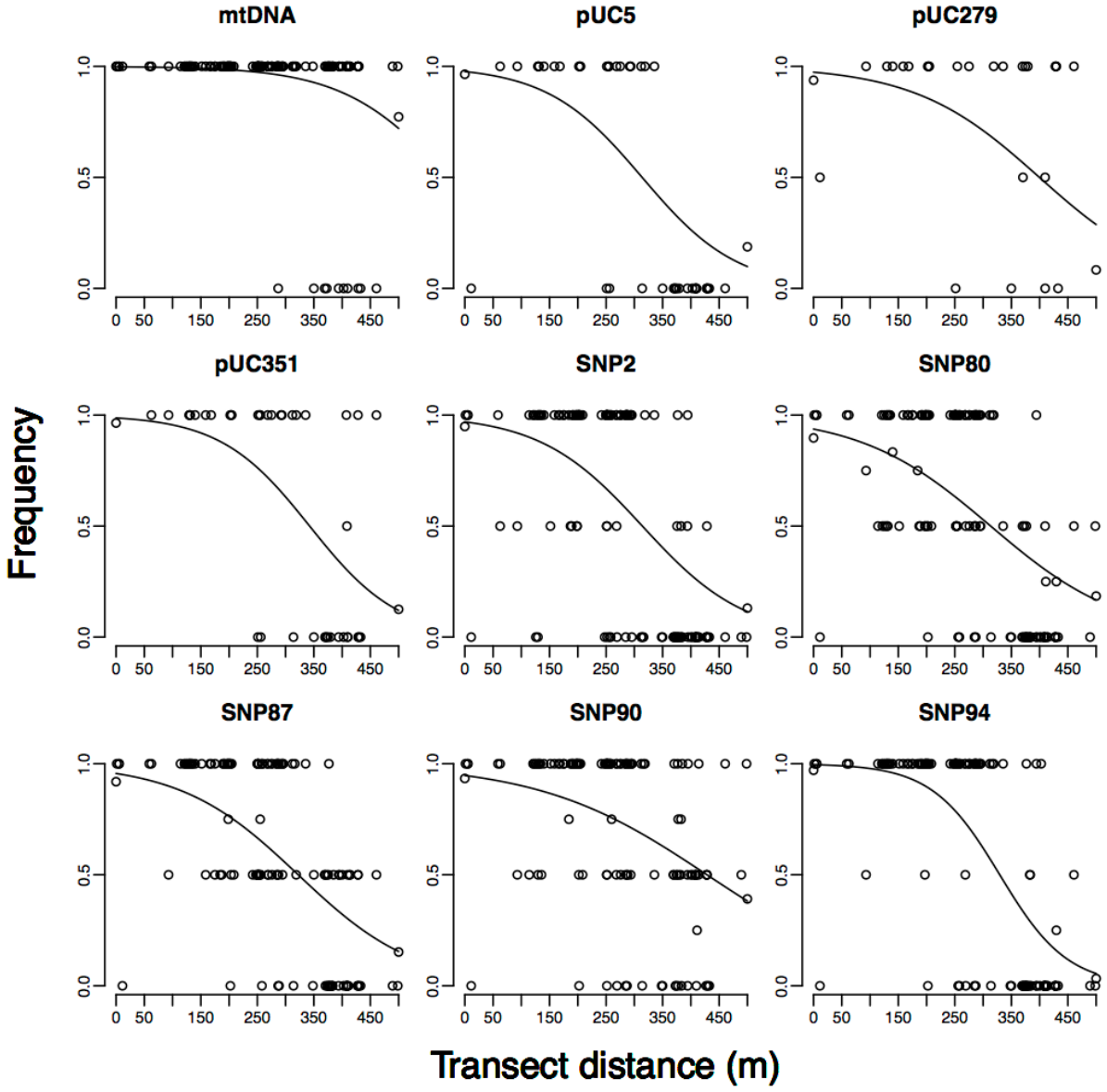


Figure S5.2 (Continued)

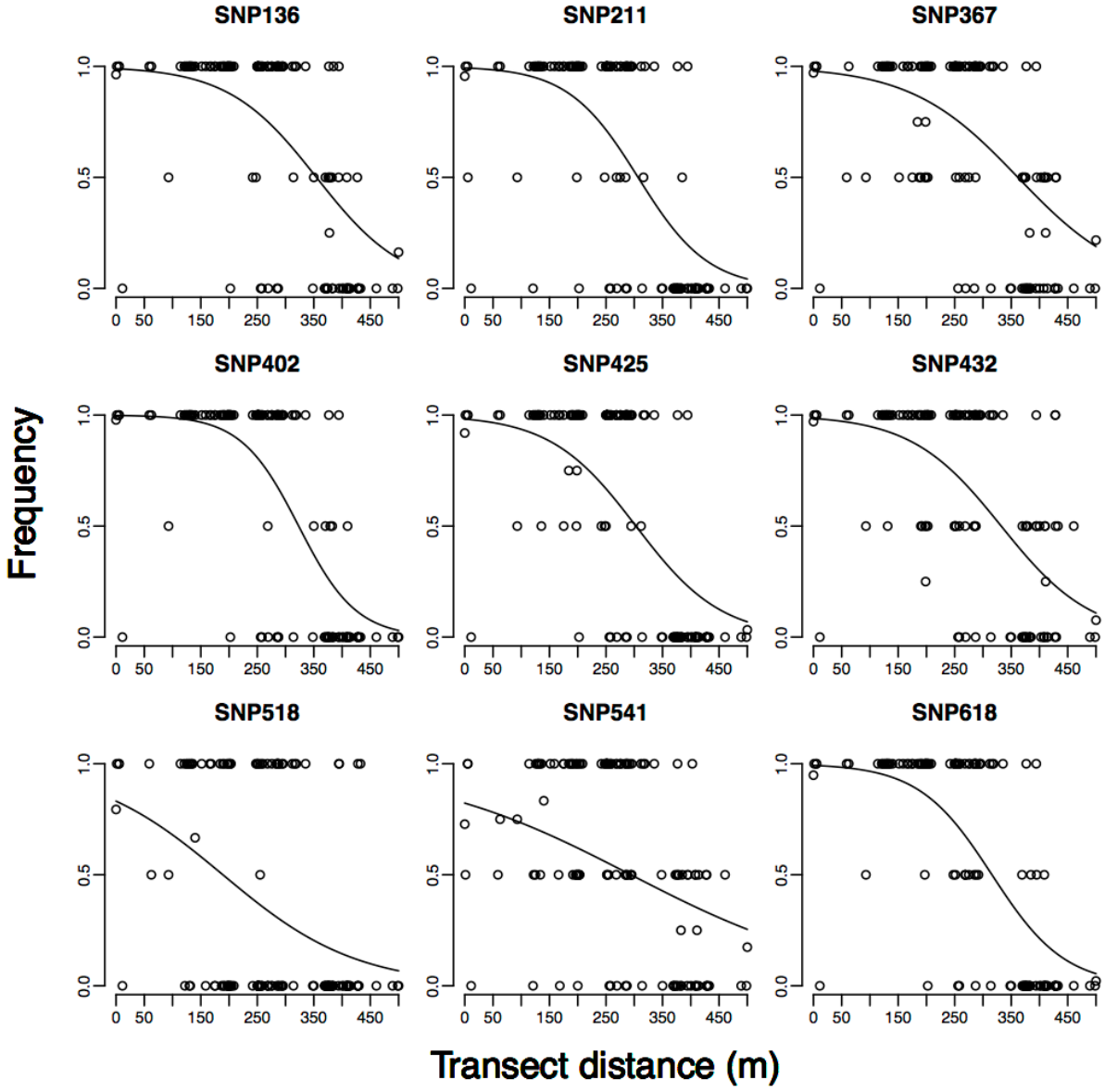


Figure S5.2 (Continued)

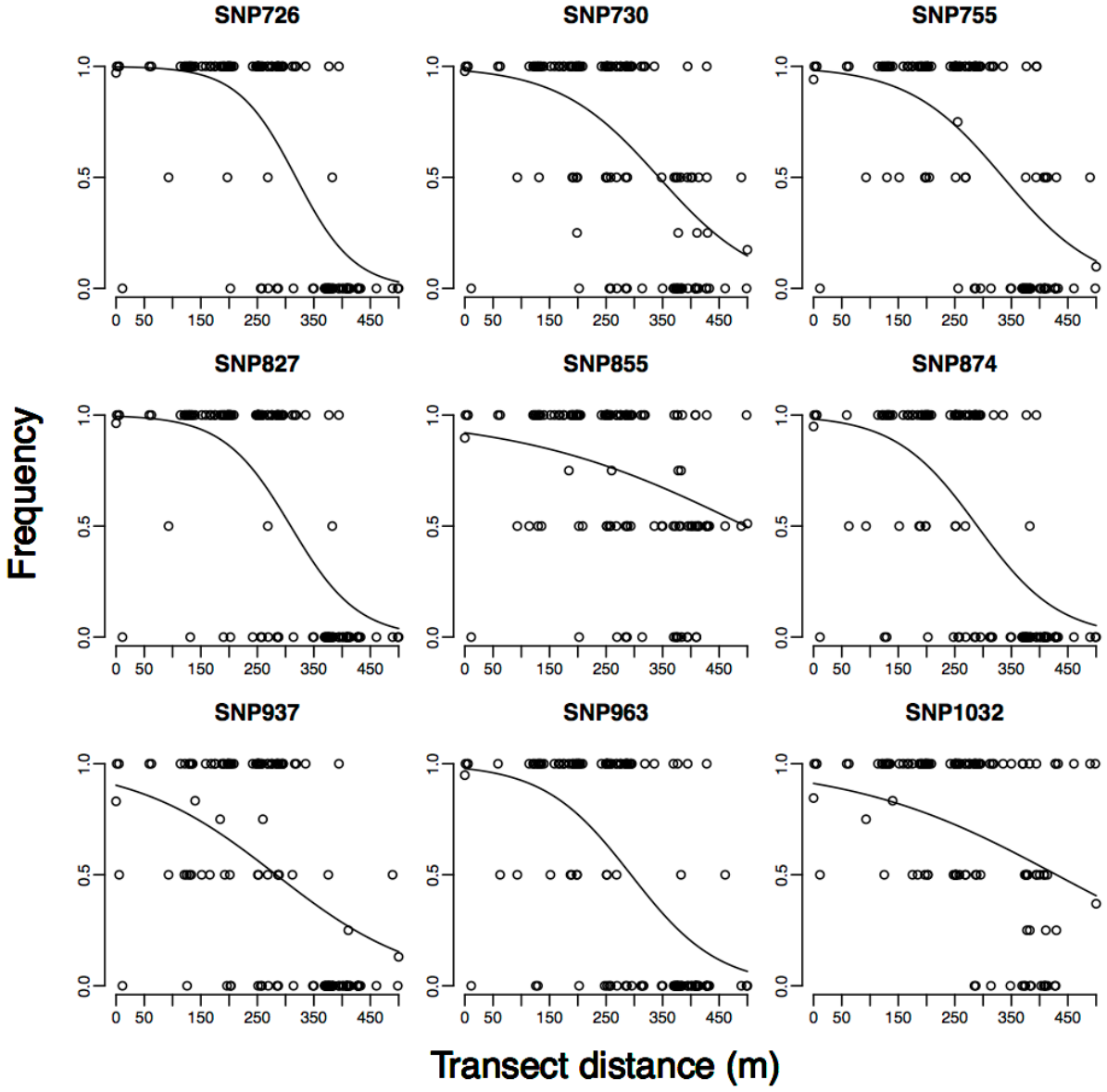


Figure S5.2 (Continued)

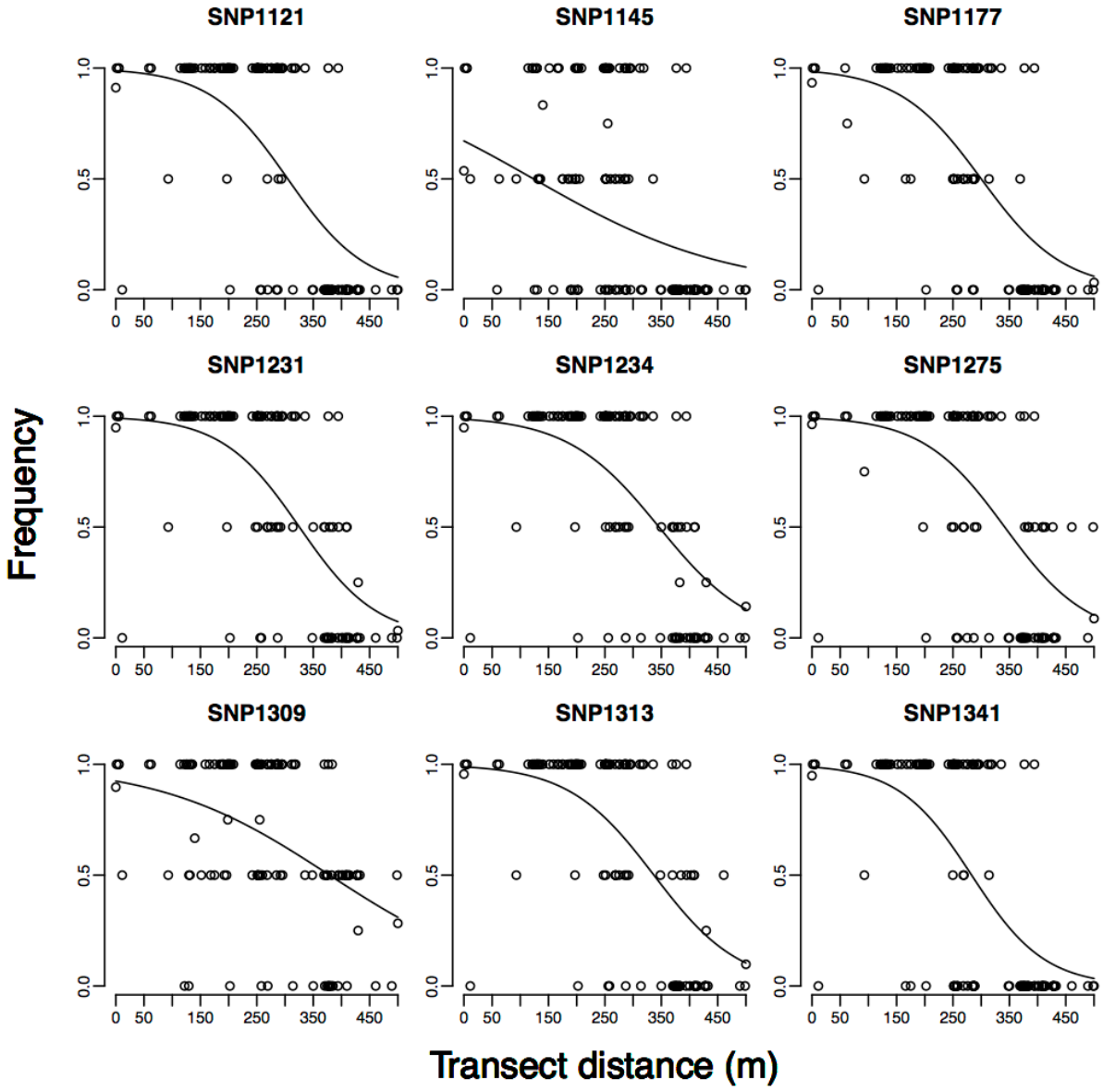


Figure S5.2 (Continued)

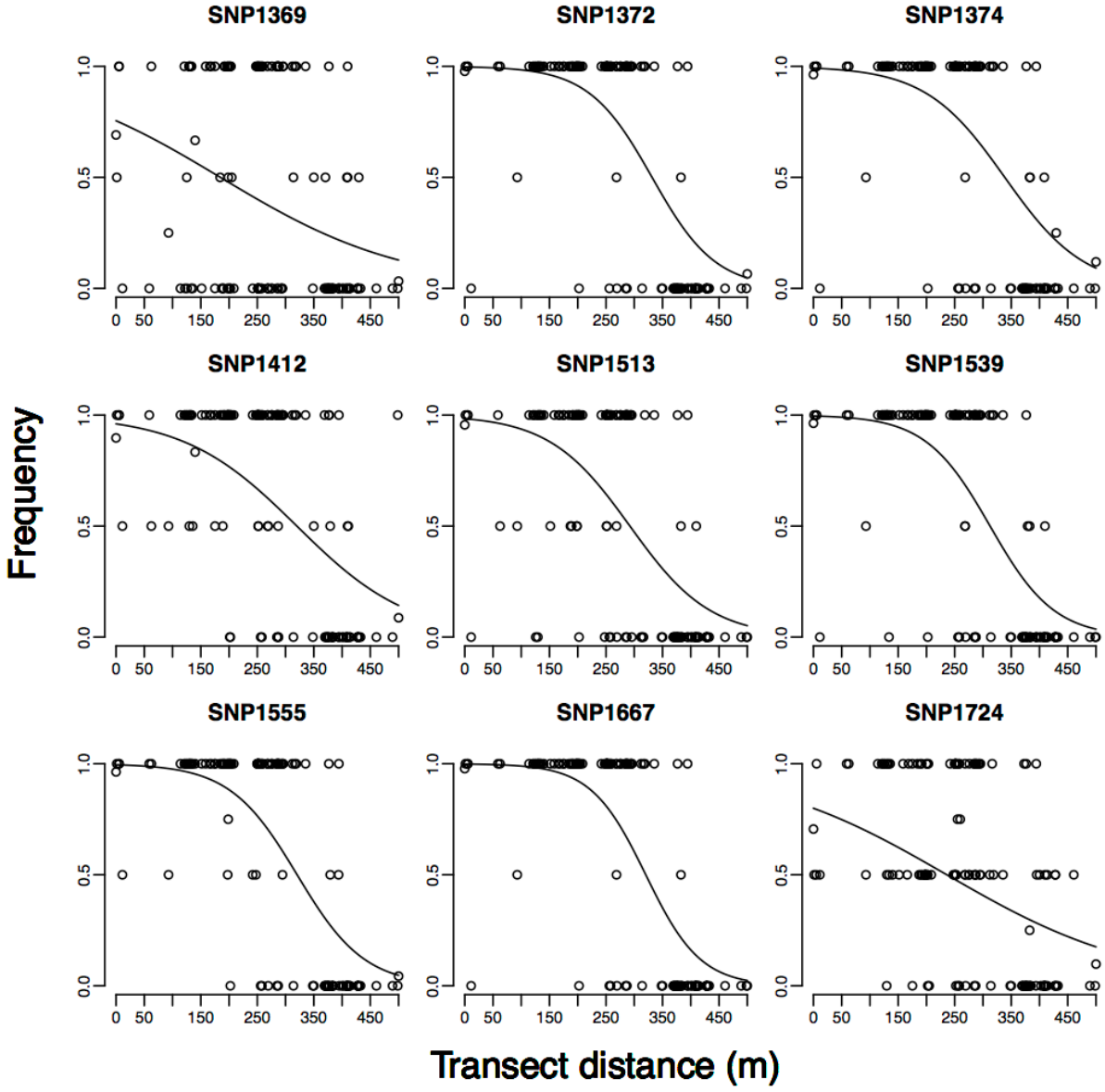


Figure S5.2 (Continued)

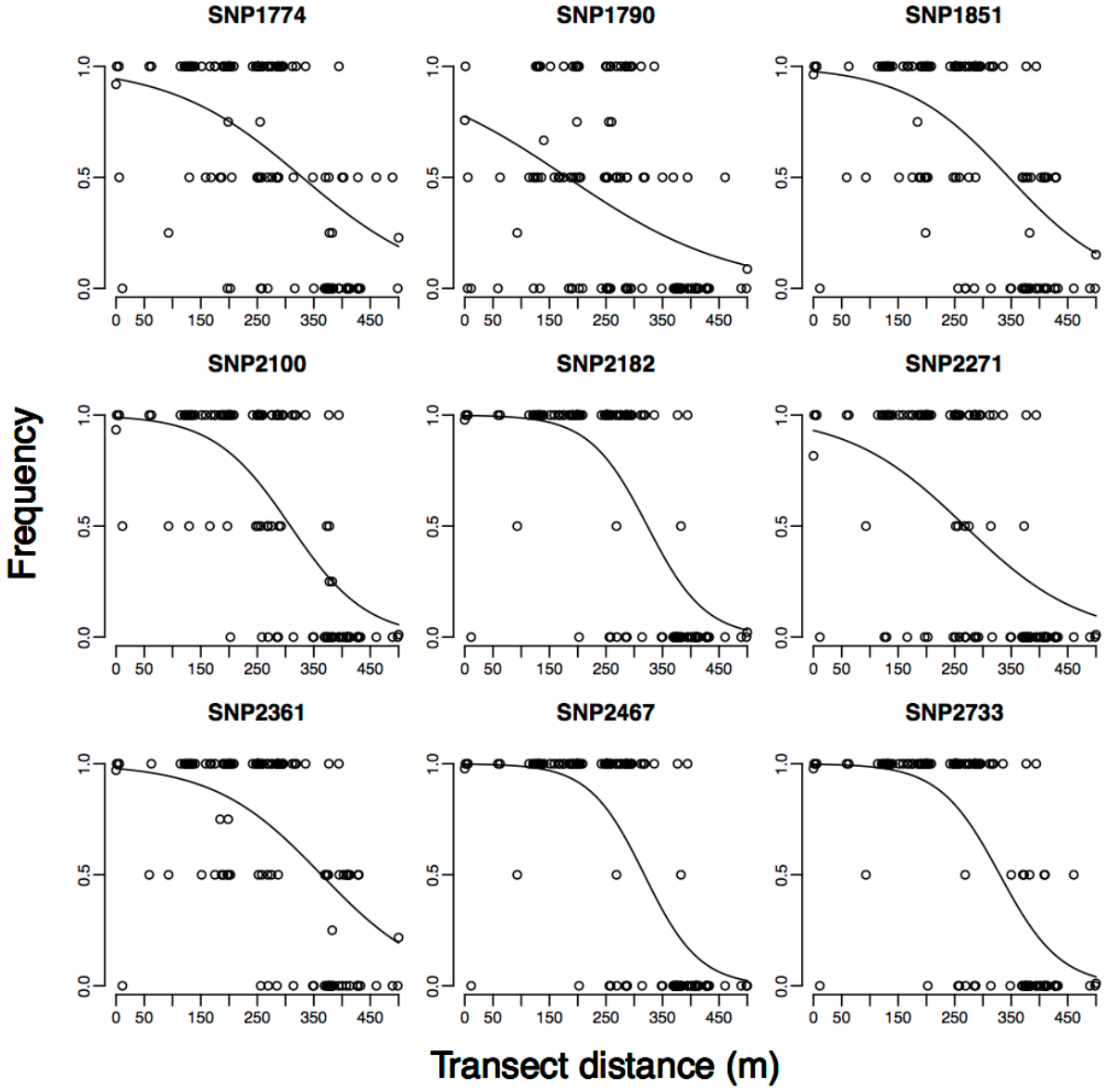


Figure S5.2 (Continued)

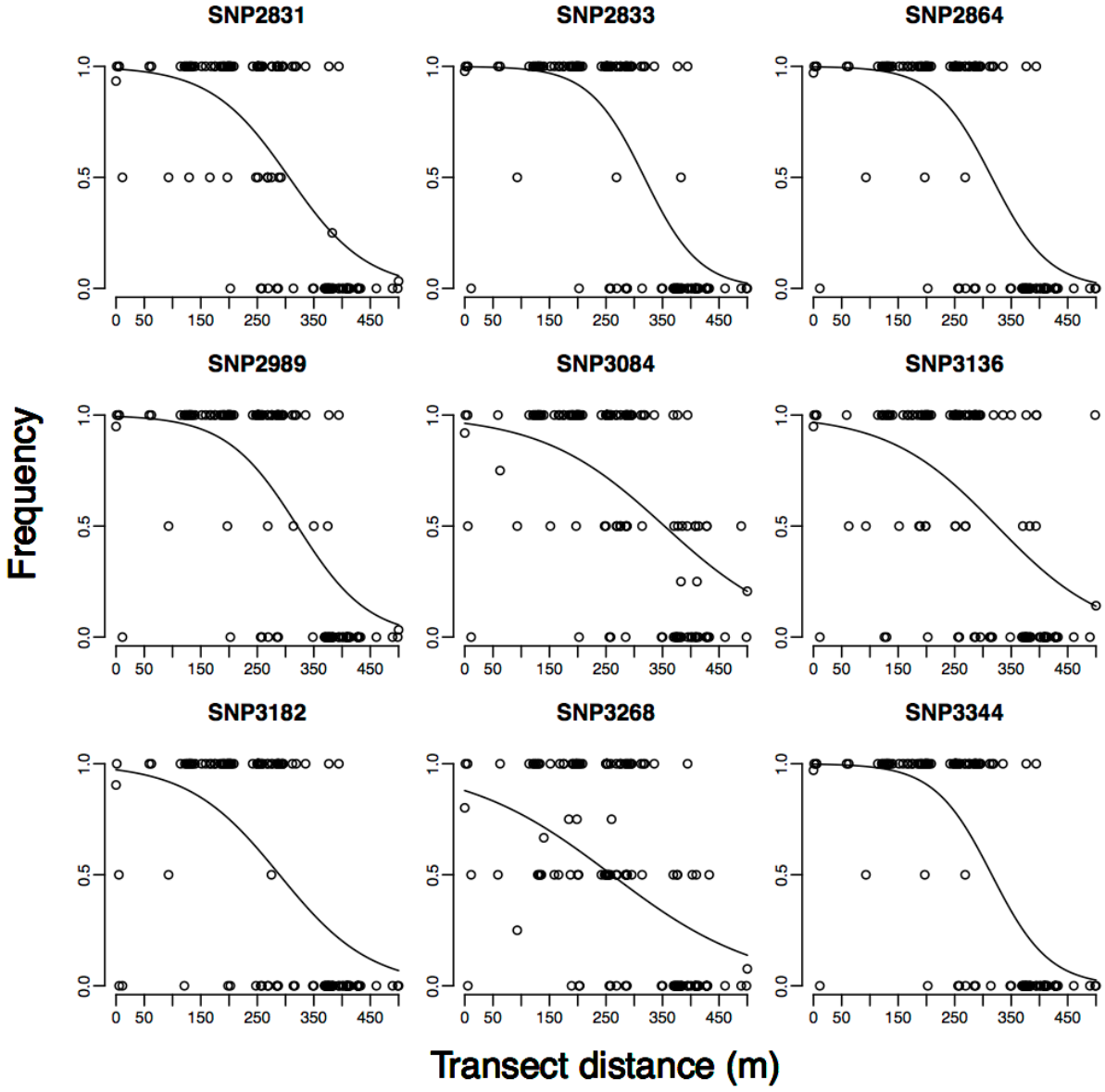


Figure S5.2 (Continued)

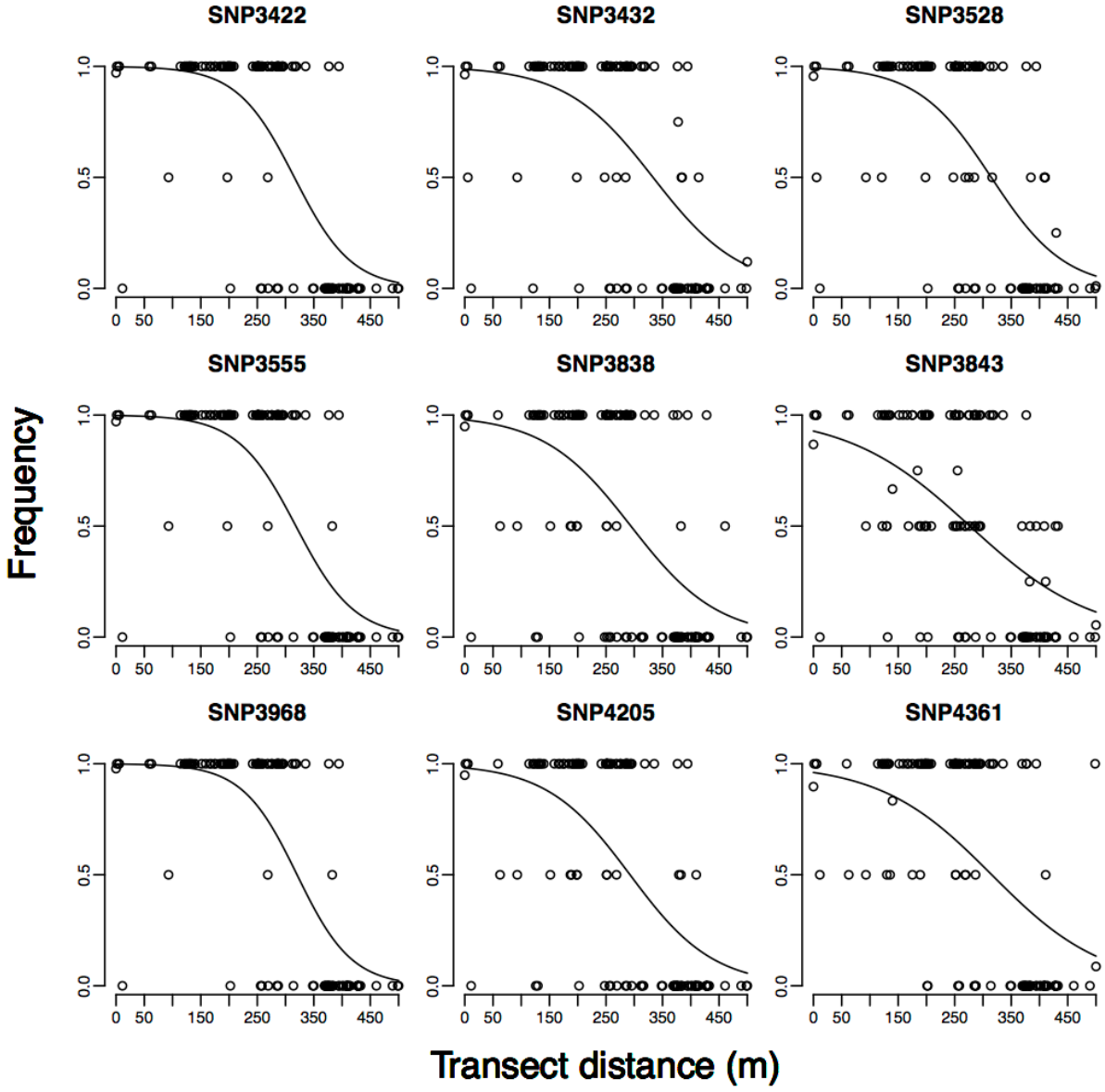


Figure S5.2 (Continued)

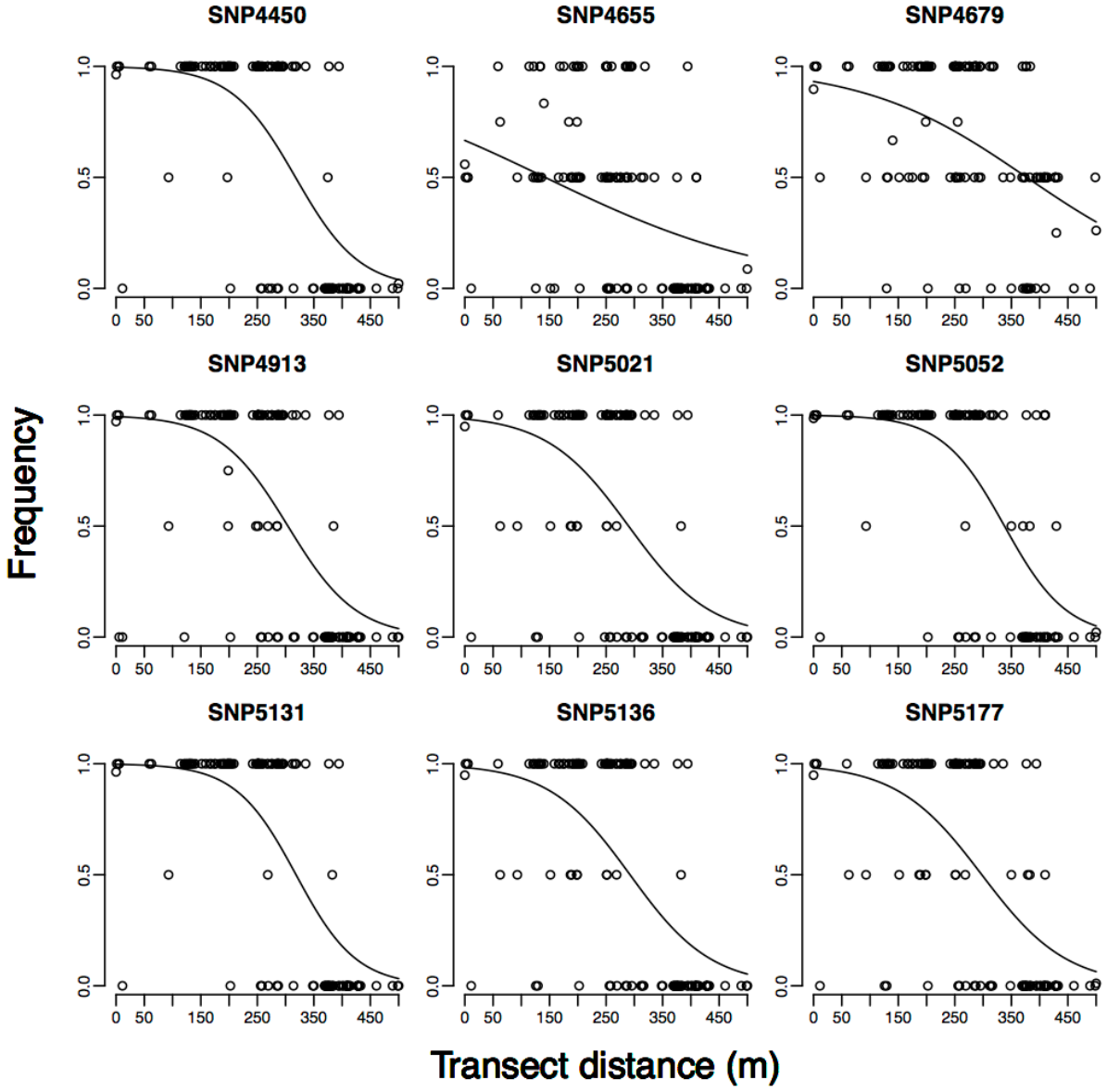


Figure S5.2 (Continued)

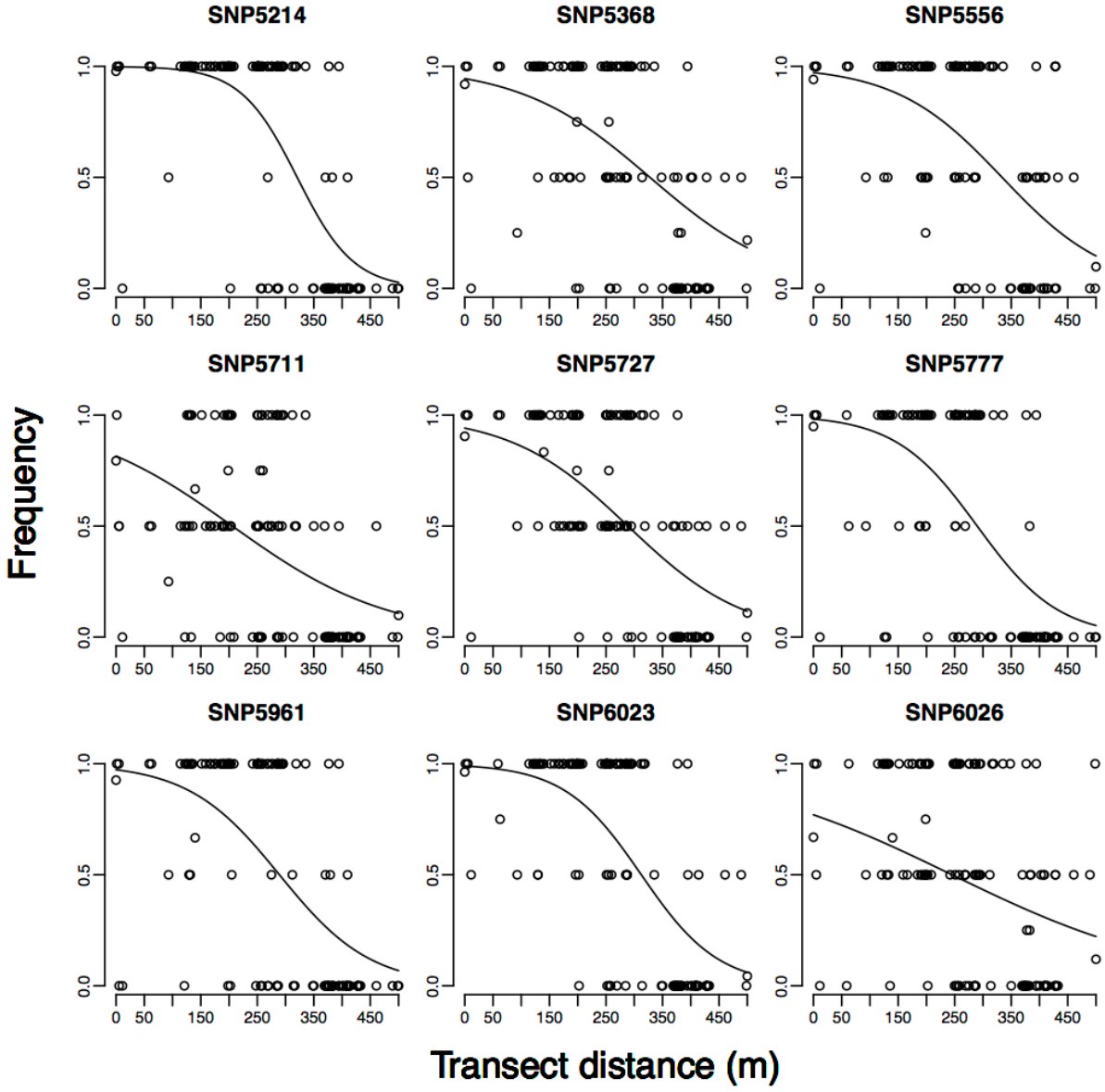


Figure S5.2 (Continued)

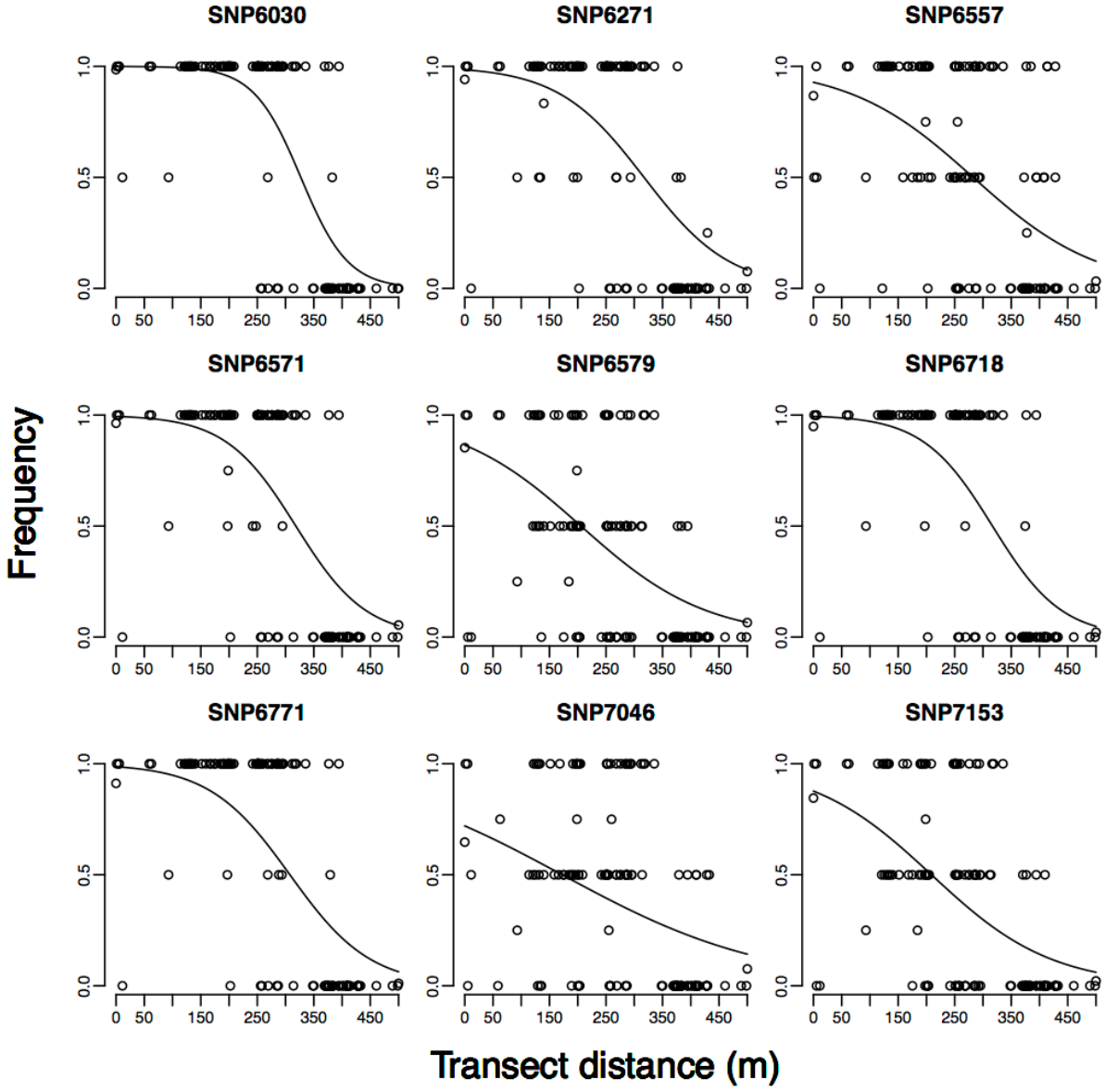


Figure S5.2 (Continued)

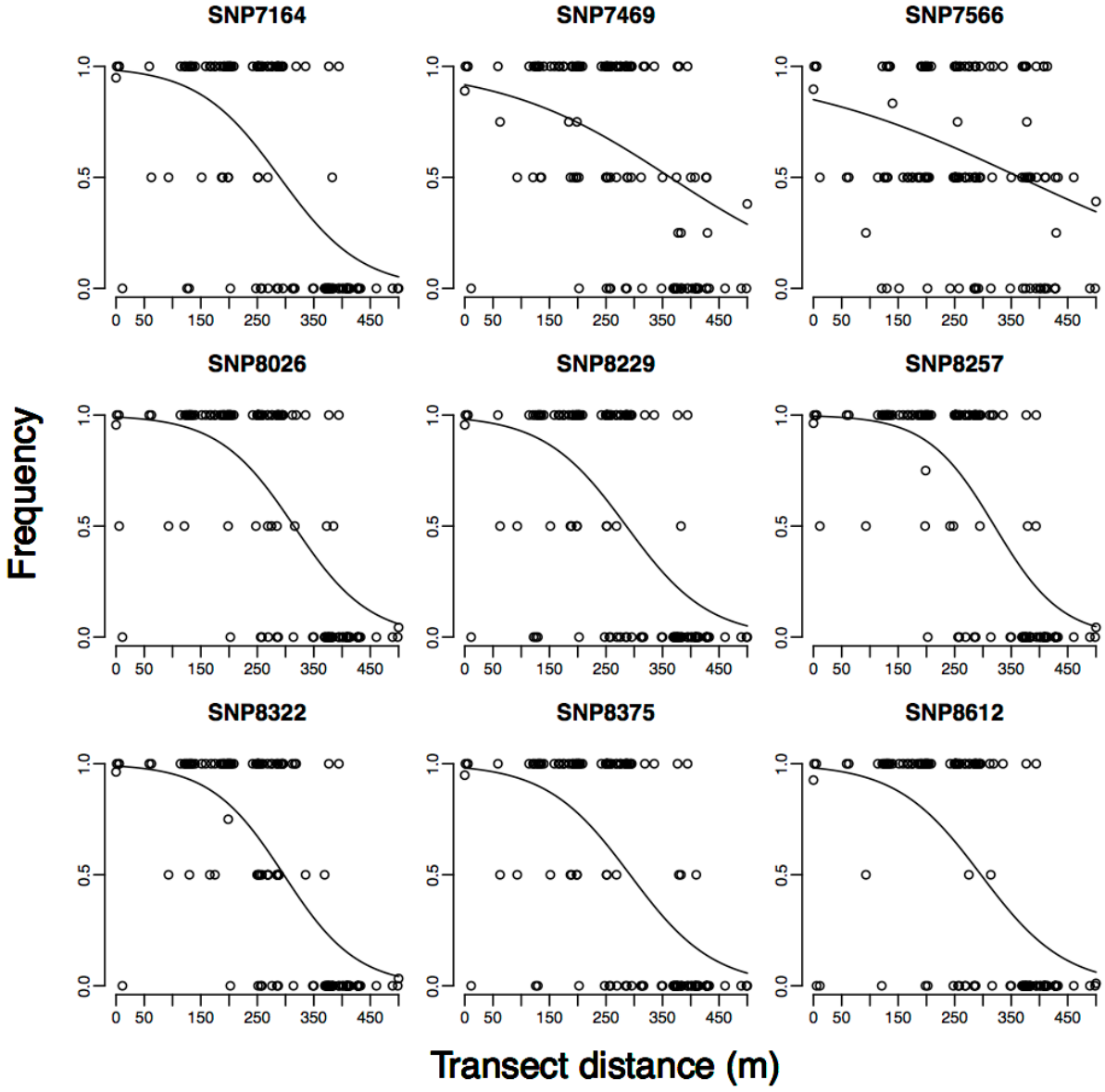


Figure S5.2 (Continued)

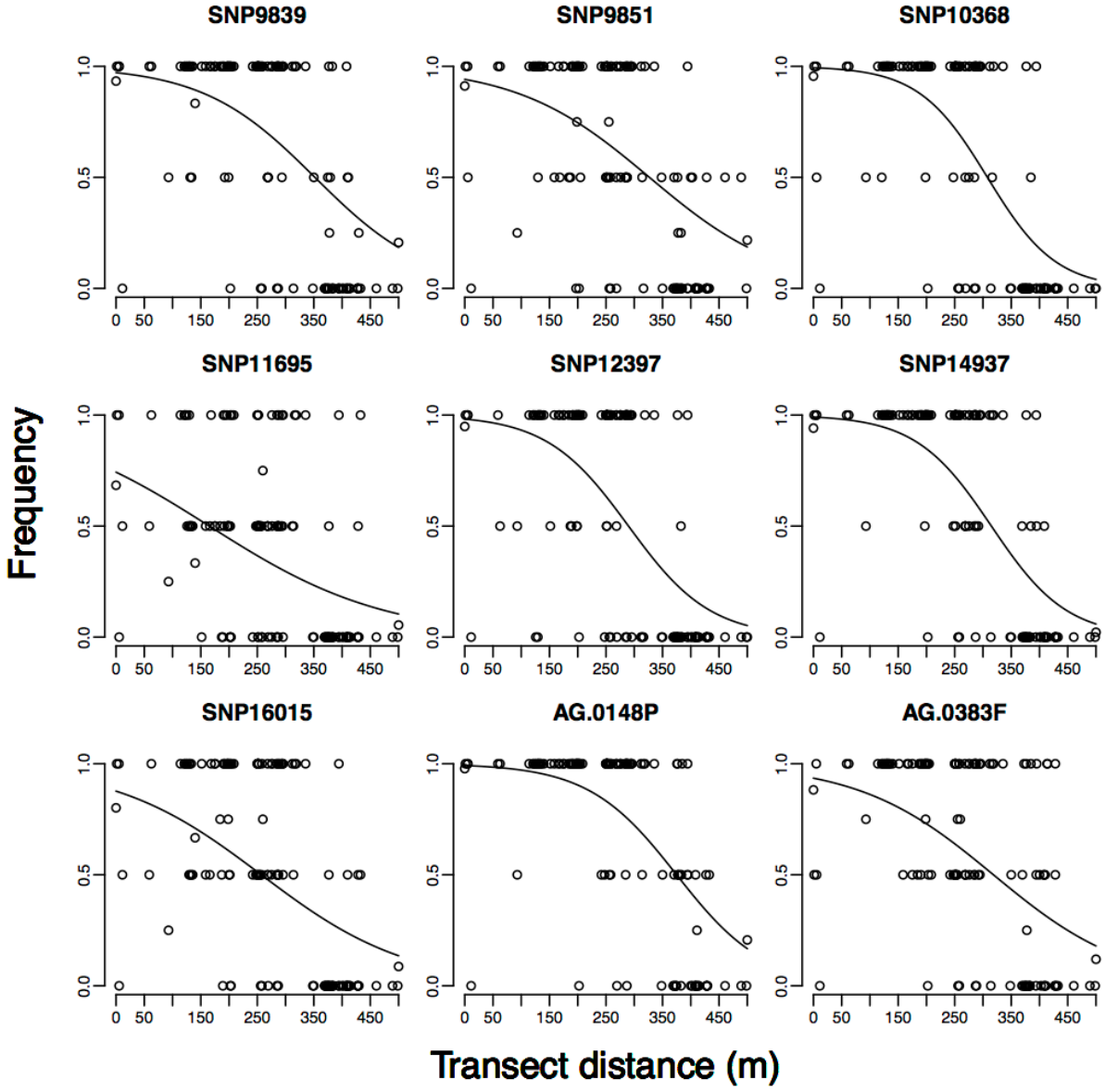


Figure S5.2 (Continued)